# Boredom and Distraction in Multiple Unmanned Vehicle Supervisory Control

M.L. Cummings[1], C. Mastracchio, K.M. Thornburg, & A. Mkrtchyan
Massachusetts Institute of Technology
Cambridge, MA

Operators currently controlling Unmanned Aerial Vehicles report significant boredom, and such systems will likely become more automated in the future. Similar problems are found in process control, commercial aviation, and medical settings. To examine the effect of boredom in such settings, a long duration low task load experiment was conducted. Three low task load levels requiring operator input every 10, 20, or 30 minutes were tested in a four-hour study using a multiple unmanned vehicle simulation environment that leverages decentralized algorithms for sometimes imperfect vehicle scheduling. Reaction times to system-generated events generally decreased across the four hours, as did participants' ability to maintain directed attention. Overall, participants spent almost half of the time in a distracted state. The top performer spent the majority of time in directed and divided attention states. Unexpectedly, the second-best participant, only 1% worse than the top performer, was distracted almost one third of the experiment, but exhibited a periodic switching strategy, allowing him to pay just enough attention to assist the automation when needed. Indeed, four of the five top performers were distracted more than one-third of the time. These findings suggest that distraction due to boring, low task load environments can be effectively managed through efficient attention switching. Future work is needed to determine optimal frequency and duration of attention state switches given various exogenous attributes, as well as individual variability. These findings have implications for the design of and personnel selection for supervisory control systems where operators monitor highly automated systems for long durations with only occasional or rare input.

---

[1] Corresponding Author, 77 Massachusetts Avenue 33-311, Cambridge, MA 02139, (617) 252-1512, (617) 253-4196 (Fax), MissyC@mit.edu

**Keywords**: boredom, distraction, multiple unmanned vehicles, human supervisory control, task load, workload


**1.    Introduction**

While automation has greatly reduced operator workload and generally enhanced safety  in supervisory control settings, humans can fall prey to "the ironies and paradoxes of automation" [1]. Increased automation can lower an operator's  task load to the point where vigilance is negatively affected and boredom possibly results. Unfortunately, as increased automation shifts controllers into system management positions, monotony, loss of vigilance, and boredom are likely to proliferate [2].

With recent advances in autonomous flight control of unmanned aerial vehicles (UAVs), it is not uncommon in search and reconnaissance missions for a UAV pilot to spend the majority of the mission waiting for a system anomaly to occur, with only occasional system interactions. For example, in describing the difficulties in flying Predator UAVs due to the few tasks requiring human input, one pilot said, "Highly skilled, highly trained people can only eat so many peanut M&Ms or Doritos or whatnot…There's the 10 percent when it goes hot, when you need to shoot to take out a high-value target. And there's the 90 percent of the time that's sheer boredom—12 hours sitting on a house trying to stay awake until someone walks out [3]." In a recent Predator operations study, 92% of pilots reported 'moderate' to 'total' boredom [4].

In these systems, automation has become so reliable at maintaining programmed system states (such as flying a holding pattern), human operators have little to do unless either the system experiences an anomaly or exogenous events occur that require mission replanning. This reduced need for interaction can result in a lack of sustained attention, leading to boredom with ultimately negative performance consequences such as missed alerts. Moreover, boredom may be a factor that induces complacency, which is also a significant concern in supervisory control systems [5].

This problem of requiring a highly trained specialist to supervise a highly automated system with relatively few required interactions that results in boredom is not unique to UAV operations, as similar problems have been reported in air traffic control settings [6], the supervision of process control plants [7], train engineers [8] anesthesiologists [9, 10] and even commercial aviation pilots [11, 12]. This problem was recently brought to the attention of the general public when two Northwest pilots overflew Minneapolis by 90 minutes because the plane was on autopilot and, as reported by the FAA, the pilots became distracted by their laptops [13], presumably because the enroute portion of the flight required so little interaction that the pilots sought stimulation from another source.

Given that automation is becoming more prevalent in such complex systems, more research is needed in predicting negative consequences as a result of long periods of inactivity and boredom, with possible deadly consequences, as well as designing to mitigate these problems. As will be discussed in more detail in the next section, while there has been significant research into the impact of fatigue as well as the vigilance decrement [14] in supervisory control domains, there has been relatively little focus on the impact of boredom on human performance, particularly in supervisory control environments [15-17].

Through a human-in-the-loop experiment with operators independently controlling multiple unmanned vehicles under low task load environments, we show that the lack of sustained attention and boredom can negatively impact performance in terms of reaction times and mission performance metrics. In addition, while generally more focused attention on task objectives resulted in better performance as compared to participants who were completely distracted, we demonstrate that participants who can effectively self-regulate their attention management can successfully perform their objectives while spending a significant amount of time distracted. Furthermore, we show that operators spend substantially less time in divided attention states as compared to focused and distracted attention states. These results suggest that technology and personnel selection interventions could be leveraged to help operators cope with low task load and resultant boredom in such supervisory control settings.

2.        **Background**

There is no commonly agreed-upon definition of boredom. Random House Dictionary [18] defines it as a state of weariness caused by dullness and tedious repetition. Various researchers have defined boredom as an unpleasant, temporary affective state resulting in a human's lack of interest for a specific current activity [17], or a decreased arousal state associated with feelings of repetitiveness and unpleasantness [19, 20]. Descriptors commonly associated with boredom include tedious and monotonous. It has been shown that boredom produces negative effects on morale, performance, and quality of work [21].

It is generally recognized in cognitive psychology that boredom has both trait (i.e., individual propensity toward boredom [22], also known as boredom proneness) and state (i.e., environmentally-driven transient feelings [23]) aspects. Ultimately, boredom is an affective and subjective state often resulting in frustration, dissatisfaction, melancholy, and distraction [24]. We define a distraction as having occurred when an operator consciously elects to divert attention from a primary task to an unrelated task (such as no longer monitoring an air traffic display in favor of talking on a cell phone.)

In addition to high job dissatisfaction, boredom has been linked to greater anxiety and stress [17, 25], particularly for military personnel [26], as well as premature death due to cardiovascular disease [27, 28]. Other long term consequences of boredom can include absenteeism and poor retention [17]. These problems directly negatively impact supervisory control domains where highly skilled operators are few and require significant training. The US Air Force is currently struggling to retain enough UAV pilots, and the boring environment is one of a number of causes for the low retention [29].

Boredom has been shown to occur in mentally demanding environments [23, 30-34]. This highlights the subjective nature of boredom, and it is important to make the distinction between task load, or the number of tasks that must be completed in the environment, and workload, which is the subjective interpretation of the task load by an individual operator. Since it is not an objective measure, boredom can

be considered a performance-shaping factor which influences vigilance, motivation, stress, and fatigue [9, 35]. However, to what degree still remains an open question.

Fatigue, typically caused by hours of continuous work or too much work, and boredom often occur concurrently [36]. However, fatigue is not a necessary and sufficient prerequisite for boredom, i.e., operators can be well rested, but still suffer performance decrements as a result of boredom. Boredom and fatigue are related in that a pre-existing lack of sleep could bring about a bored feeling earlier, or that a boring work environment induces fatigue. However, limiting shift work time in a supervisory control UAV environment proved to be a poor safeguard as even a four-hour work shift still resulted in fatigue and boredom [4]. This paper focuses on strictly boredom, which is not exacerbated by any pre-existing fatigue factors.

In terms of the impact of boredom on supervisory control performance, previous studies on air traffic control monitoring tasks showed that participants reporting high boredom were more likely to have slower reaction times and worse performance than participants reporting low boredom [37, 38]. Similarly, participants who reported higher subjective, task-related boredom also had slower reaction times. Furthermore, a study of U.S. air traffic controllers showed that a high percentage of system errors due to controller planning judgments or attention lapses occurred under low traffic complexity conditions [39].

Boredom is intrinsically related to vigilance, which is defined as "a state of readiness to detect and respond to certain small changes occurring at random time intervals in the environment" [40]. Given that vigilance tasks are, by definition, repetitive, it is not surprising that research has shown that participants in some laboratory vigilance experiments consider such tasks to be boring and monotonous [41-43]. This is not to say all vigilance tasks are under-stimulating, as recent research has highlighted that vigilance tasks can be demanding [44]. While the existence and negative impact of the vigilance decrement (i.e., a performance decline over time for vigilance tasks) in supervisory control settings is well established [45-47], relatively little research has been conducted to assess the impact of resultant boredom in such settings, particularly in terms of operator coping strategies.

5

Laboratory studies attempting to measure negative impacts of the vigilance decrement and boredom are difficult due to the time scale needed (hours), participant recruitment and retention, and the subjective nature of boredom, i.e., not all participants are guaranteed to actually become bored. Many vigilance studies have been conducted in strict laboratory environments with far more stimulus events than are realistic [48]. Vigilance studies typically focus on repetitive tasks, often with short response stimulus intervals on the order of seconds or just a few minutes [49, 50]. Such studies are representative for machine-paced and continuous control environments such as airport baggage inspection tasks and train engineers who must engage a "dead man's switch" every 30-50s, and have generally shown that such environments can reduce physiological arousal and lead to boredom in most people [51-55].  More recently, several researchers have attempted to use neurophysiological assessment techniques such as fMRI [56] and galvanic skin response [57] to measure displeasure and boredom, but such work is in its early stages and there are no commonly-accepted psychophysiological measures of boredom.

Studies that require near-continuous human-system interaction, such as those environments typical in monotony studies, do not generalize well to supervisory control domains where operators are expected to monitor and intermittently interact with an automated system, with potentially long periods of inactivity. Thus, there is a need to conduct more targeted research in environments where there is literally almost nothing to do [17], which, as evidenced by the earlier UAV pilot quotes, is the case in some highly automated unmanned vehicle systems where persistent surveillance is a primary mission objective. Our attempt to address this gap specifically for UAV environments, and more generally for highly automated supervisory control environments, is detailed in the next section.

## 3.    Method

As will be detailed in the following sections, we hypothesized that given a four hour long study, with approximately only one new task to attend to per hour, that participants would not be able to effectively sustain attention after 20-30 minutes, in keeping with the vigilance literature [15]. Moreover,

we hypothesized that participants would become bored, resulting in poor focused attention and continued degraded performance. We also hypothesized that time spent distracted over time would increase, and those participants that were the most distracted would perform the worst. Lastly, this experiment was also an observational one since we wanted to determine how operators coped with the low task loading and what interactions and distractions might occur as a result.

**3.1 Apparatus**

This experiment employed a collaborative multiple unmanned vehicle simulation environment that leverages decentralized algorithms for vehicle routing and task allocation. For this experiment, participants were responsible for one rotary-wing UAV, one fixed wing UAV, one Unmanned Surface Vehicle (USV), and a Weaponized Unmanned Aerial Vehicle (WUAV). The objective of the simulation was to find as many targets as possible, and destroy the hostile ones over a four-hour test session. This time was selected since the only other UAV study that examined the effects of boredom [4] was also a four-hour study.

The UAVs and USV were responsible for searching for targets. Once a target was found through panning and zooming across an image sent from the vehicle, the operator identified the target given the icon presented (blue rectangle was friendly, yellow clover was unknown, red diamond was hostile). Chat messages from a virtual commander instructed participants what priority level to assign to targets (i.e., high, medium, or low). High priority targets were to be destroyed immediately, low priority targets could be ignored if needed, and medium target prosecution was left to the discretion of the operator. Neutral targets would be designated as friendly or hostile later in the experiment through a chat message, while one or more vehicles tracked hostile targets until the human operator approved WUAV missile launches.

While operators do not specify any vehicle parameters in terms of altitude, speed, specific headings, etc., they perform higher-level tasks like creating search tasks, which dictate areas on the map where the operator wants the unmanned vehicles (UxVs) to specifically search. Operators also have scheduling tasks, but these are performed in collaboration with the automation. The autonomous planner

recommends schedules, but the operator can accept, reject, or modify these plans. This autonomous planner only communicates to the vehicles a prioritized task list, and the vehicles sort out the actual assignments among themselves by "bidding" on the tasks they estimate they can accomplish, manifested through a consensus-based auction algorithm [58]. In the course of this market-based auction scheme, vehicles bid on tasks while attempting to minimize the revisit times between tasks.

However, if the operator is unhappy with the UxV-determined search patterns or schedule, he or she can create new search tasks, in effect forcing the consensus-based decentralized algorithms to reallocate the UxVs. This human-automation interaction scheme is one of high level *goal-based* control, as opposed to more low-level *vehicle-based* control.

Participants interacted with the simulation via two displays. The primary interface is a map display (Figure 1a). The map shows both geo-spatial and temporal mission information (i.e., a timeline of mission significant events), and supports an instant messaging communication tool. This chat box simulates a command center that provides high-level direction and intelligence about targets in the designated area.

In the map interface, operators identify targets and approve weapon launches and insert new search tasks as desired or dictated via the chat box. The performance plot in the lower left corner gives operators insight into the automated planner performance as the graph shows expected performance (given an *a priori* cost function) compared to actual performance. When the automation generates a new plan with a predicted performance better than the current plan via a decentralized, polynomial-time, market-based algorithm [58], the Replan button turns green and flashes. This illuminated button indicates that a new plan is ready for approval, and when selected, the operator is taken to the Schedule Comparison Tool (SCT), detailed next (Figure 1b).
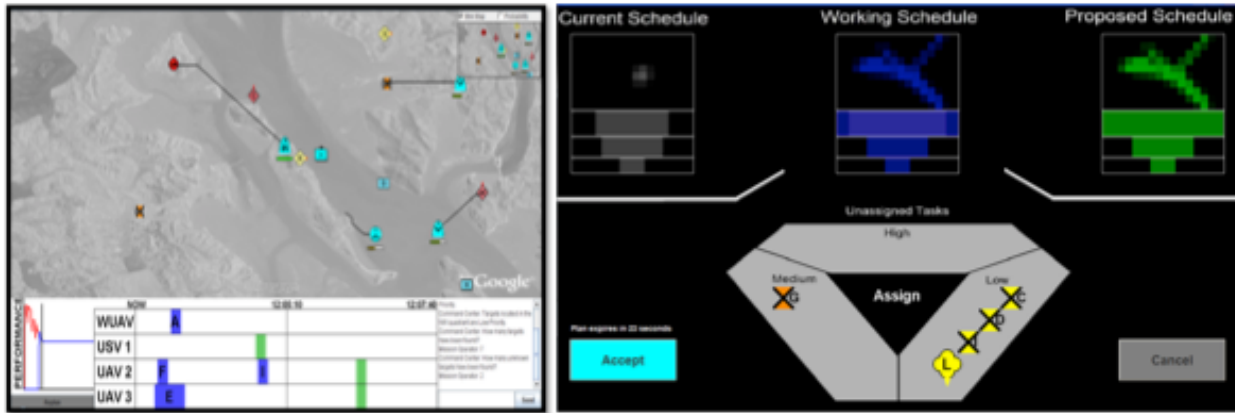
**Figure 1a: Map display (left) and 1b: Schedule Comparison Tool (right).**

Clicking on the Replan button in the lower left corner of the map display allows operators to view the Schedule Comparison Tool (SCT) (Figure 1b) anytime they desire. The three geometrical forms at the top of this display are configural displays that leverage direct-perception interaction [59] to allow the operator to quickly compare three schedules: the current, working, and proposed schedules. The left form (gray) is the current UxV schedule. The right form (green) is the latest automation proposed schedule. The middle schedule (blue) is the working schedule that results from the user modifying the plan by querying the automation to assign particular tasks. The rectangular grid on the upper half of each shape represents the estimated area that the UxVs will search according to the plan. The hierarchical priority ladders show the percentage of tasks assigned in high, medium, and low priority levels.

When the operator first enters the SCT, the working schedule is identical to the proposed schedule. The operator can conduct a "what if" query process by dragging the desired unassigned tasks into the large center triangle. This query forces the automation to generate a new plan if possible, which becomes the working schedule, causing the configural display to change accordingly. Ultimately, the operator either accepts the working or proposed schedule or can cancel to continue with the current schedule. Details of the interface design and usability testing can be found in [60]. This environment is both a computer simulation and supports actual flight and ground capabilities [61]; all the decision support displays discussed in this paper have operated actual small air and ground unmanned vehicles.

9

**3.2 Participants and Procedure**

The population for this long duration study consisted of 30 paid participants with 19 men and 11 women. Ages ranged from 19 to 32 years (Mean (M) = 23 years, Standard Deviation (SD) = 3 years). Forty-three percent of participants had prior military or Reserve Officers' Training Corps (ROTC) experience, but none had any multiple unmanned vehicle control experience (as no such operators presently exist). All participants reported good to great sleep for the previous two nights. Ten test sessions were conducted with three participants each within a mock command and control center. Each participant was responsible for his or her own simulation, i.e., there were no dependencies in a test session between the three participants. Participants were tested three at a time since unmanned vehicle operating environments typically contain multiple personnel responsible for dissimilar tasks.

Each participant had two computer monitors that displayed either of the screens shown in Figure 1, and another that was used at the participants' discretion. Participants interacted with this system and other computer functions through use of a standard keyboard and mouse. All wore headphones that sounded alerts when chat messages arrived, when new targets were found by the UxVs, and when they were prompted to replan. Participants were compensated $125 for their participation and also informed that the individual with the highest performance score would receive an additional $250 in the form of a gift card.

The experiment was conducted on a Dell Optiplex GX280 with a Pentium 4 processor and an Appian Jeronimo Pro 4-Port graphics card. Each display's resolution was 1280x1024 pixels. In order to familiarize each participant with the interfaces in Figure 1, a self-paced, slide-based tutorial was provided, as well as a ten-minute familiarization session with the experimenter. Then each participant conducted the control task for approximately an hour at a moderate task load level (they were required to interact with the system approximately half of the time) to ensure they were experts with the system. By the end of this training, participants had to demonstrate to the experimenter that they could effectively use the SCT, including understanding that they did not always have to follow the automation's recommendations. All

participants exhibited these characteristics by the 30-minute training mark. After the practice session, participants were informed how well they did in terms of the percentage of targets they successfully destroyed, but intentionally were given no feedback in terms of strategies.

Participants were not prohibited from engaging in tasks unrelated to the UxV control task. Checking email, reading books, talking, and eating were allowed, though these activities were neither encouraged nor discouraged. One goal of this study was to determine what types and how often participants would engage in self-selected distractions. Refreshments were provided to the participants, and the same food varieties were made available to all participants. Testing started at mid-day for every test session. The test administrators remained in an adjacent room and came into the test room four separate times, at the same time, to query if any of the participants needed to use the restroom or were experiencing any problems.

The primary goals for each participant included finding moving targets in the environment and destroying hostile targets in an expedited and efficient manner. In this four-hour session, only four targets were available to be found and of those four, only two were designated hostile. These targets were released at a rate of one per hour, but this does not necessarily mean they were found at this same rate. If operators did not pay attention and maintain oversight of where vehicles traveled over time, it was possible to not find any targets over the four hours. It took almost an hour for a search vehicle to move from one side of the map to the other, reflecting a typical persistent surveillance environment. The participants were not told this information.

Performance dependent variables included the number of targets found and number of hostile targets destroyed. The primary measure of objective workload was determined through utilization, i.e. the time a participant was engaged in one or more system-generated tasks, divided by the total time available. Utilization has been used to estimate objective workload in human performance models [62, 63] and supervisory control empirical studies have shown utilization to be a reliable predictor of high workload [64-66].

The following utilization metrics were explored in this study: (1) required utilization, or the percentage of mission time the operator spent performing mandatory tasks required by the system; (2) self-imposed utilization, or the percentage of mission time the operator spent doing mission tasks not required, such as adding search tasks he or she felt would improve the automation's performance, and (3) total utilization, the sum of required and self-imposed utilization. In terms of utilization, operators were considered "busy" for mission-related tasks when performing one or more of the following tasks: creating search tasks, identifying and designating targets, approving weapons launches, interacting via the chat box, and replanning in the SCT.

Monitoring time was not included in percent busy time since participants were not explicitly trained to scan the display or perform any actions when the system provided no tasking. It is possible that during these times, occasionally participants could be cognitively engaged in the task, but given that we could not absolutely determine this, we limited the busy classification to observable, actionable behaviors. A self-rated busyness 5-point Likert metric was also collected as a subjective measure of workload since the objective workload measure focused on operator percent busy time. In this way, we attempted to more closely align the subjective measure of workload with the objective measure of task load, which also focused on how busy operators were. This scale was chosen since humans generally can only make subjective absolute assessments across five categories [67].

This experiment was a between-subjects design, defined in terms of three different replan intervals. A replan interval is the time interval separating automated requests for operator help in evaluating the computer-generated schedule. The three intervals were 10, 20, and 30 minutes, so operators were prompted to consider a new schedule 6, 3, or 2 times per hours. At these intervals, each operator was prompted by the green illumination of the replan button along with an aural replan alert, and then presented with a proposed automated schedule in the SCT. All aural alerts were designed to cue the participants to look at the screen.

These three conditions were selected since they represent the low extreme of task loading for this interface and were empirically determined via a discrete event simulation model [66, 68], targeting 5, 10, and 15% utilization levels. By comparison, a previous moderate workload study with this same system prompted operators every 0.5s - 120 *seconds* [68]. The intervals were counterbalanced and randomized across participants. So for each group of 3 participants, not everyone experienced the same interval in order to preserve ANOVA independence statistical assumptions. Participants' interactions with the system were automatically recorded. In addition, participants were knowingly videotaped via wall-mounted cameras to track their attention throughout the test sessions. While there is debate as to whether such recording has no effect or distorts results [69], all participants experienced the same exposure and we relied on statistical analysis to mitigate any negative effects.

### 4.      Results and Discussion

A one way analysis of variance (ANOVA) was used to assess differences across the various dependent variables for parametric data and Kruskal-Wallis tests were used for non-parametric data that violated normality and homoscedastic assumptions, $\alpha = 0.05$. It should be noted that while there are similarities in this experimental setting and other supervisory control environments like air traffic control, these results are unique to this environment. Moreover, the results should be interpreted in light of the relatively small sample size of 30 participants.

### 4.1 Mission Performance

To assess how well the participants performed across the three different replanning rates of 10, 20, and 30 minutes, a one way ANOVA was conducted on two scores: 1) the Target Finding Score (TFS), and 2) the Hostile Destruction Score (HDS). These scores, assigned individually to each participant on a 0-1 scale with 1 representing best performance, encapsulate the primary objectives expressed to the participants. TFS represented the number of targets found in the mission, weighted by how many total

could be found and by how quickly they were found, so that targets found more quickly resulted in a higher score. Similarly, the HDS score was the sum of hostile targets destroyed weighted by the maximum number that could be destroyed and also accounted for how quickly the hostile targets were destroyed. Note that TFS and HDS data from three participants were removed because of data loss and corruption.

There were no statistical differences across the three replanning intervals for both of these performance scores, meaning that regardless of how often the automation asked an operator for help, operator performance was unaffected. One might expect boredom to decrease, and thus performance to increase with increasing replanning intervals, however, in this study, even the 10 minute interval was not stimulating enough to make a difference. Current research is underway to determine more optimal replanning intervals, both in terms of human and algorithm performance.

To place the overall performance of this low task load group in context, their TFS and HDS scores were approximately one-fourth of those of a similar study that used the exact same test bed with the exact same training, but under moderate task load conditions (adjusting for the differences in the length of the experiments between the moderate and low task load studies (10 minutes versus 4 hours) as well as number of available targets (10 vs. 4)) [70]. Not only did this low task load group, on average, perform substantially worse than those in a moderate task load setting, but 5 participants in this low task load study did not destroy any hostile targets at all.

**4.2 Utilization**

Various aspects of utilization, measures of objective workload, were investigated due to the unique nature of this goal-based control paradigm. While the automation's replanning rate could be varied experimentally to control how often the automation *asked* for help (i.e., required utilization), operators could replan anytime they wished by adding search tasks to the list of vehicle tasks, thereby *preempting* the automation and intervening when the system did not ask for help (i.e., self-imposed utilization).

One experimental confound in such a study that reflects the stochasticity of the real world is that even though the automation replanning rates were determined prior to the experiments based on predictions from a simulation model, how much utilization was required for each test session was directly dependent on whether participants found the targets, and how much each participant elected to preempt
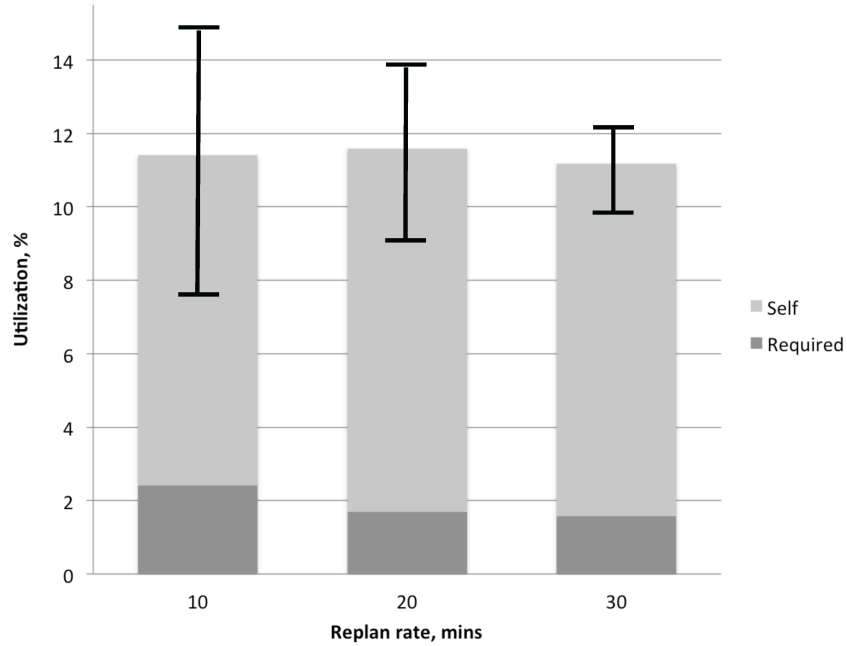


**Figure 2: Mean Utilization by Replan Interval, with 95% Confidence Intervals for Total Utilization**

the automation. Figure 2 demonstrates the utilization metrics (required, self-imposed, and total utilization) across the three different replanning intervals, with several interesting results.

First, there was statistically different required utilization across the three different replanning intervals. (Kruskal-Wallis $\chi2 = 13.224$, $p = 0.001$). The 10 minute interval group had an average of 2.46% required utilization (SD = 0.49%), the 20 minute interval group had an average of 1.69% required utilization (SD = 0.15%), and the 30 minute interval group had an average of 1.63% required utilization (SD = 0.34%). Operators were tasked more to help the automation in the 10 minute replanning condition as compared to the other two lower task load conditions. However, it should be noted that overall,

operators across all three conditions were only required to interact with the system roughly 2% of the entire mission time, or 5 minutes.

Self-imposed utilization (operators electing to add search tasks and replan when the automation did not require it) is effectively a measure of how busy operators would make themselves in such a low task load environment, beyond what was required. The underlying algorithm, a consensus-based bundle market-based algorithm, was not guaranteed to produce optimal results so operator intervention could, in fact, improve the overall solutions [58].

The results in Figure 2 demonstrate that regardless of how often the automation asked for help, participants gravitated toward the same self-generated task load (Kruskal-Wallis $\chi 2 = 0.439$, $p = 0.803$). This led to statistically no difference across total utilization (the summation of self-imposed and required utilization), as seen in Figure 2. Because the standard deviations were so low for required utilization (in all cases, $< 0.5\%$), the 95% confidence interval for total utilization is effectively the same for self-generated utilization. So, regardless of the fact that some participants were tasked more than others, they all gravitated to the same narrow range of total utilization with an average of 11.0% (SD = 3.36%), approximately 26 minutes of the entire four hour test session.

In a post-experiment survey, participants were also asked to rate how busy they felt they were over the four hour test session at the conclusion of the experiment (1 = idle through 5= extremely busy). Across the three replanning sessions (10, 20, 30 mins), the median scores were all 2 and not statistically different using the Kruskal Wallis test. Participants generally did not feel they were overly taxed, which agrees with the overall nearly uniform total utilizations in Figure 2. In fact, no participants rated themselves as very or extremely busy (4 and 5 on the Likert scale) and only 2 participants out of 30 rated themselves as Busy (3). Moreover, all participants rated themselves as confident, very confident, or extremely confident about the correctness of their actions.

**4.3 Reaction Times**

Due to the low task load and the randomness that each participant could inject into the simulation, there were few repeatable comparable events that could indicate whether reaction times were negatively influenced, which is expected given the vigilance literature. However, there were three such possible repeated measures throughout the experiment: 1) Reaction times to automation prompting to consider a replan, 2) Reactions to text messages asking for information, and 3) Prompts from the system to generate search tasks. An audio alert was sounded for all three tasks, in addition to visual cues.
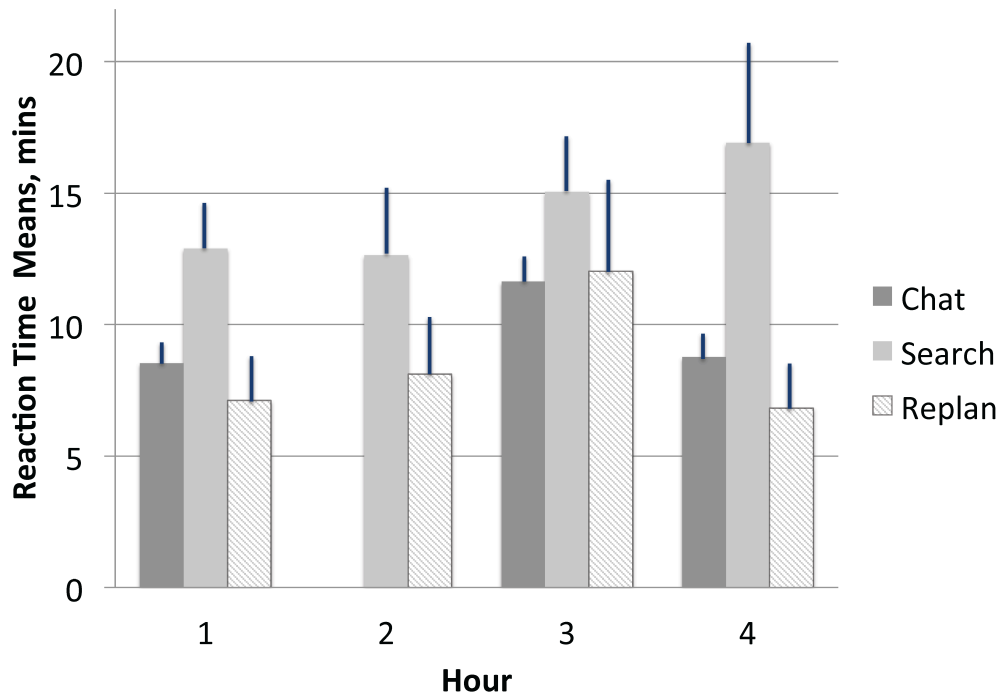


**Figure 3: Chat, Search, and Replan Task Reaction Times with One Standard Error Shown.**

Figure 3 illustrates an observed increase in reaction times for all three events across the four hour experiment. There was no statistical difference across the three replanning levels given a multivariate ANOVA, so the data were combined in Figure 3. It should also be noted that there were no chat messages in the second hour, thus no data. Of note is the improvement in the chat and replan times in the last hour of the experiment, when we hypothesize participants became more engaged since they knew they were in the final hour of the experiment, as evidenced by conversations between participants to this end.

However, the search task creation times continued to erode and by the conclusion of the experiment, were a full 4 seconds slower than they were at the start. This could be an indication that participants considered the search creation tasks to be the lowest priority of the tasks and not substantially beneficial to the system's overall performance.

Gender has been suggested as a contributor to individual variability of perceived boredom [18, 80, 81]. However, there were no statistically significant effects of gender on reaction times, or any other dependent variable.

## Attention States

Given that participants had significant opportunity for distraction in this study, we measured where their attention was directed at any given time and when they switched attention. We used three attention categories in this study: 1) Directed, which is when participants were directing their gaze at the interface or interacting with the interface, (2) Divided, when participants were looking or glancing at the
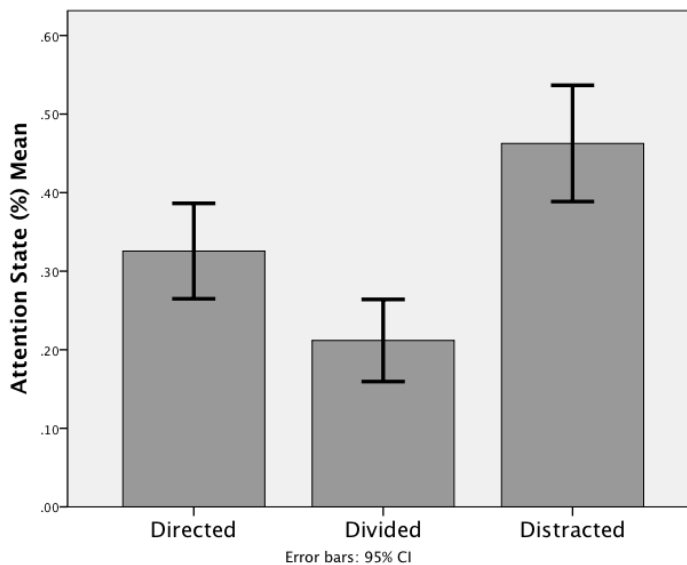


**Figure 4: Aggregate Attention States with 95% Confidence Interval**

interface but also engaged in other tasks such as talking to other participants, eating while watching the screen, etc., and (3) Distracted, which was coded as a participant not in a physical position to see the

interface, such as turned around in a chair while talking to other participants, at the table getting something to eat, working on a personal laptop, etc.

These categories are based primarily on physical actions, and coded from videos taken of the entire experiments, a technique used in coding attention states in previous studies [71, 72]. We recognize that they are not absolute in that just because someone was looking at a screen, this does not mean he or she was cognitively engaged with a UAV control activity. However, these states were unambiguous to code and given that this study was designed to reflect actual command and control center environments, these states replicate the body positions that supervisors in command centers such as Air Traffic Control centers use to judge whether personnel are engaged in their tasks. Two personnel coded the four-hour long videos, and a third quality assurance person randomly checked ten percent of all results and if any discrepancies were discovered, the entire video was coded again. The inter-rater agreement (Cohen's Kappa) was 85%.

In Figure 4, we collapsed the three replanning rates into a single category since a multivariate ANOVA revealed no significant differences for attention states across the three replanning rates. The results of Figure 4 demonstrate that across the four hours, participants spent most of their time in a distracted state (45%). This is of obvious importance since, by our definition, operators could not see the interface at all when distracted. While all participants had headphones to warn them of new events, in many cases participants ignored the alarms, either intentionally or due to their source of distraction. Occasionally participants removed the headphones even though they were instructed to wear them at all times. It is not clear whether the presence of the aural cue influenced the time spent in divided or distracted states through possibly engendering inappropriate trust [73, 74]. However, such event-based aural cues are standard in the design in supervisory control domains.

The most common sources of distraction noted by the video coders in the coding of attention states were talking with other participants in the room, accessing the Internet, using cell phones, and eating snacks. The distraction behaviors seen in this study are not unlike others seen in similar

19

environments. Anesthesiologists reportedly read, listen to music, talk on the telephone, or converse with their colleagues during periods of low workload [9].

While participants generally spent approximately one-third of the time directing their gaze towards the interface or interacting with the system (i.e., utilization), another interesting result is how little time participants actually spent in a divided attention state (~21%). These empirical results add to an increasing body of literature that shows that people are not as effective at multitasking as they might think [75, 76], and that when given the opportunity, people typically gravitate towards a single attention state (in our case, either directed or distracted.) These results have important implications for supervisory control tasks, as previous research has shown that monitoring performance can be improved through dividing attention across tasks [77, 78].

No attention state variables were significantly correlated with performance, e.g., there was no observable relationship between amount of distraction and performance. As will be discussed in the next section, this was primarily due to nonlinear relationships in attention management. In terms of
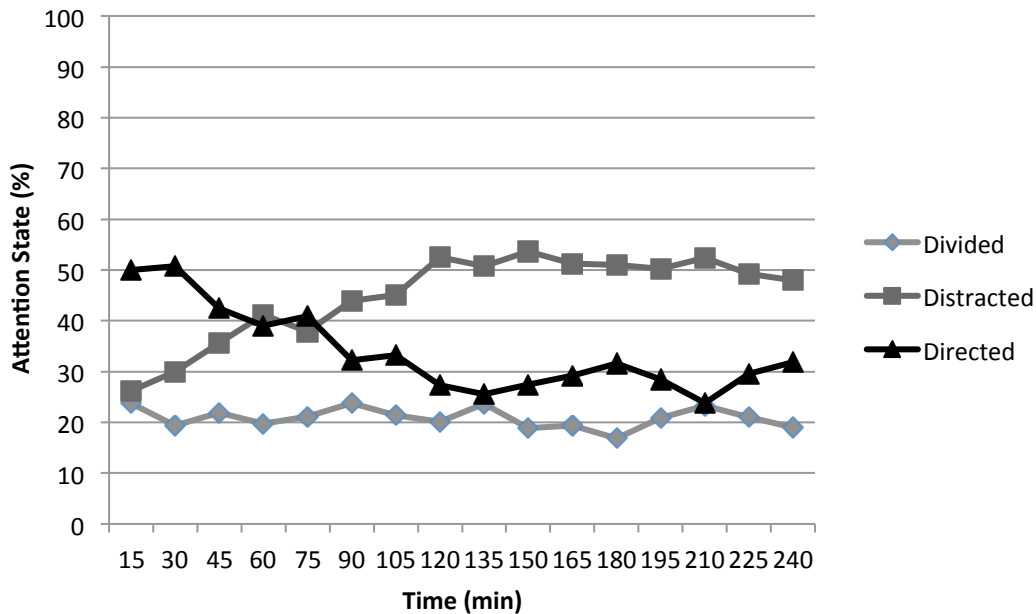


**Figure 5: Attention Management as a Function of Time.**

demographic influence assessed through a questionnaire prior to the experiment (e.g., gender, military experience, etc. The entire survey can be found in [69]), the only variable that was significantly associated with performance was the degree of video game playing. Those participants who played video games more than once weekly were statistically more likely to perform worse in this setting that those who occasionally or rarely played video games ($\rho$ = -0.304, p = 0.05). This is in direct opposition to the previous moderate workload study with this exact same interface where gamers statistically performed better [68]. Thus it appears that gaming experience can be a handicap when the exogenous event arrival pace is perceived to be under stimulating.

While Figure 4 provides important aggregate information, Figure 5 demonstrates over time (in 15 minute intervals) how participants managed their attention in this low task load setting. In Figure 5, the average standard deviation for directed attention across the 4 hour period with 15%, for divided it was 13%, and for distracted attention, 18%. Distraction was the absorbing state, and in keeping with well-established vigilance research, directed attention began a significant decline at approximately 30 minutes and at about the 1 hour interval, distraction began to outweigh directed attention, leveling off at ~50% at the 2 hour mark. Divided attention was relatively steady state at ~20% for the entire experiment. These results have important implications for supervisory control of highly automated systems, particularly those that are safety critical and may require rapid response for low probability events after long periods of operator idleness. For example, most process control operators never see the combination of events like those that precipitated the 2003 Northeast blackout that affected 55 million people, and operator lack of recognition of what was unfolding was a causal factor in the blackout [79].

## 4.4 Best versus Worst Performers

Examining the best and worst performers, in terms of their combined target finding and hostile destruction scores, provides additional insight into how attention management strategies could lead to above or below average performance. Figure 6 illustrates that, expectedly, the worst performer was
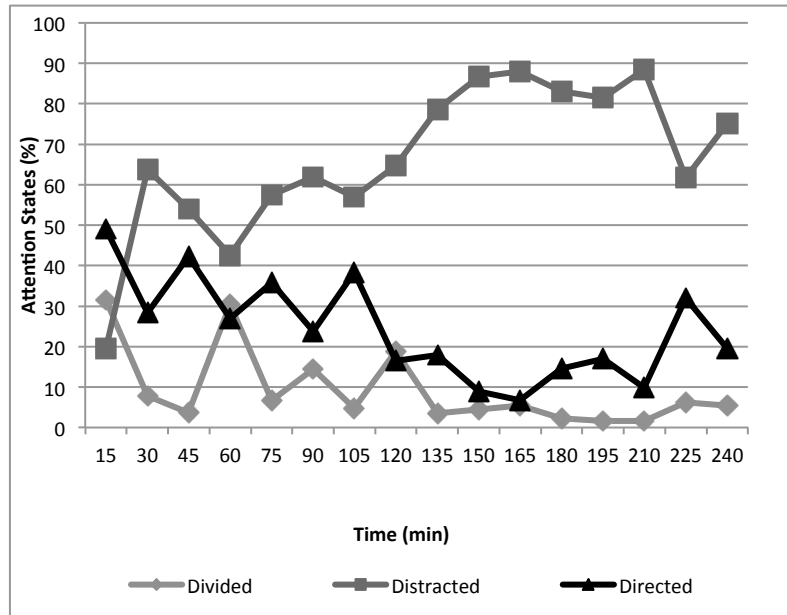
**Figure 6: Attention Profile for the Worst Performer.**

predominantly distracted throughout the entire four-hour period (66%), and spent more than an hour

~85% distracted. The worst performer, -1.9 σ for performance score, found two targets but did not destroy

any hostile targets. This participant rated himself a 2 on a scale of 1-5 in terms of busyness.

   While the worst participant's performance in Figure 6 was representative of other performers who

did poorly, there were two very different strategies exhibited by the best performers, a group of five

participants who all scored more than one standard deviation above the mean and represent the top 15%

of the sample. The best performer's strategy was to oscillate between directed and divided attention

(Figure 7a), with very little time spent in a distracted state. The second best performer also exhibited

oscillatory behavior (Figure 7b), but instead of oscillating between directed and divided, this performer

demonstrated remarkable periodicity between directed and distracted attention, with very little time in a

divided attention state. There was no correlation with this participant's attention state and aural alerts for

replanning, in that he did switch his attention when an alert sounded but also regularly switched his

attention with no cueing. Indeed, he took off his headphones for portions of the experiment.
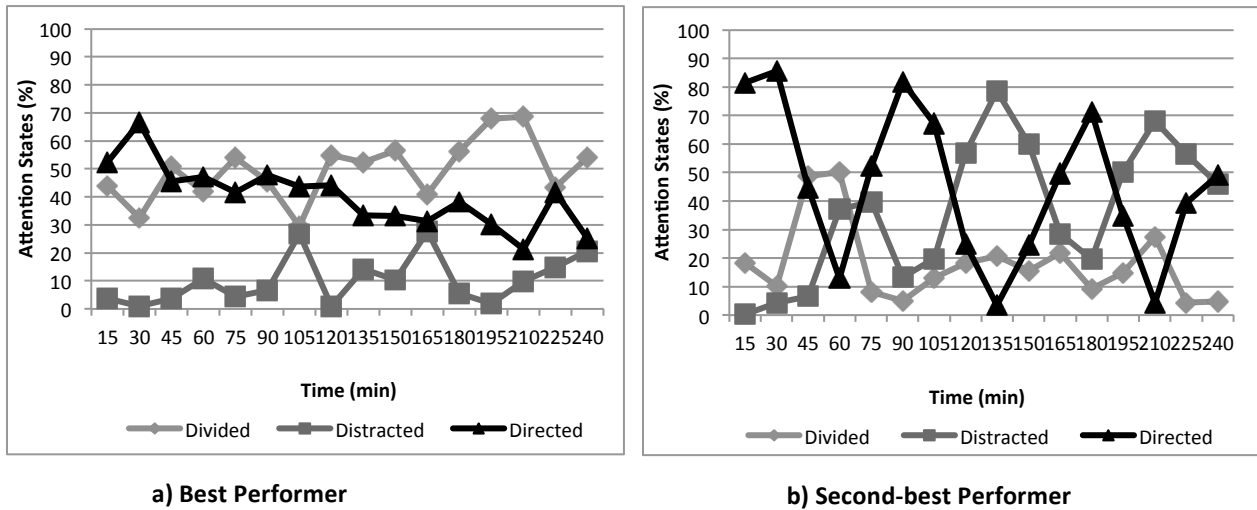
a) Best Performer                                      b) Second-best Performer

**Figure 7: Attention Profiles of the Two Best Performers.**

The remaining top three participants' attention profiles (~38% distracted) were similar to the

second best performer (~37% distracted), compared to ~10% distraction of the top performer. Also of

note is the influence of the number of years of military experience. While there was no statistically

significance for years of military experience on performance overall, four out the five top performers did

have some military experience, which could speak to the ability to more readily follow procedures and

orders.

While the best performer of Fig. 7a exhibited the most constructive coping strategies of all

participants, which generally was repeatedly creating search tasks and checking the SCT often to possibly

generate a better solution, the remaining top performers chose a completely different non-task related

coping strategy. Instead of interacting with the system frequently, they constructively interacted *just*

*enough* with the system to finish just behind the top performer.

While distraction in the human factors literature is generally seen as a problem in supervisory

control domains and was an issue for the poor performers in this low task load study, the second-best

performer was distracted for more than one-third of the mission (37%), as were the other top three

performers (38%). These operators automatically shifted their attention in a highly regular manner

between distraction and effective guidance of the automation, which we hypothesize mitigated the

negative effects of low task load and resulting boredom. Poor performers either spent too much time distracted or could not effectively regulate their attention, either in frequency or timing. Whether these issues can be addressed through personnel selection or system design are areas of future research, and are discussed in the next section.

5.      **Conclusions**

In both current and future UAV operations, as well as in many other highly automated supervisory control settings like process control, long periods can elapse without requiring any direct or indirect input from operators. Such passive monitoring, often referred to as "babysitting the automation", can be particularly difficult for operators who experience difficulties in sustaining attention and increased stress due to the tedious nature of the environment. Operators in such low task load environments are often easily distracted, but the cost and possible benefits of such distractions are not well understood. This research effort demonstrated that in a low task load setting, four distracted operators with inherent excellent attention management strategies did almost as well as the focused top performer, so the assumption that distractions will cause performance degradation may not always be accurate.

Another interesting result was that participants gravitated to the same level of self-imposed utilization, which raises more questions such as whether there is a minimum level of activity operators will seek regardless of external tasking? It is possible that such interactions are indicative of a lack of trust and acceptance of the automation. Whether such levels of self-imposed utilization would be maintained over longer and repeated experiments or decrease due to complacency or inappropriate trust is still an open question.

This research raises further questions in terms of staffing, in that several agencies would like to reduce the number of personnel in a control room, typically from a cost perspective, and often to a single person [80, 81]. This research shows that the presence of others in low task load environments may provide both negative and beneficial distraction. Previous social facilitation research has shown that the

presence of others can (but not always) increase physiological arousal with possible modest gains in simple performance tasks [82]. Moreover, organizational attempts to enforce a distraction-free or "sterile" environment may only exacerbate negative consequences of a boring environment. Banning radio listening and conversations, or limiting breaks has been shown to contribute to boredom [17, 19, 83], despite the fact that some of these secondary tasks can provide clear benefit, i.e., listening to music in a visual task can help to maintain sustained attention [84, 85].

In addition to organizational influences, a person's interest or motivation in assigned tasks also likely has an impact on individual boredom [17, 86]. We leave the question of motivation and boredom in supervisory control domains to future work, but it is a critical one that raises recruitment and retention issues already acute for the US Air Force's UAV workforce [29]. Other areas that deserve further scrutiny include the impact of training on boredom coping strategies in terms of how could or should operators be taught to sustain attention, and whether distractions add to or mitigate fatigue.

Because of the high risk involved in the operation of nuclear power plants, air traffic control, and other safety-critical supervisory control domains, personnel selection is of practical concern so identifying those personnel who can manage long stretches of inactivity in such settings is important. To this end, experience, age, intellectual capacity, gender, and personality type have all been suggested as contributors to individual variability of perceived boredom [17, 87, 88]. Moreover, in air traffic control tasks similar to the UAV control tasks in this study, it has been suggested that task characteristics of repetitiveness and traffic density may interact with individual traits (e.g., personality, experience, age) in a way that causes monotony and boredom [16]. The only clear evidence of such influences in this work was video game experience that suggested poorer performance for those with significant gaming experience, although the sample size and homogeneity of participants likely was a factor in determining other possible associations. Clearly more data is needed in these areas to understand the interaction of the individual with tedious supervisory control environments and associated distractions, and how personnel selection practices might benefit.

Ultimately, we must accept that highly automated environments will likely be considered boring environments that can and will lead to distraction. Thus the question should not necessarily be how to stop distraction but how to manage it through either personnel selection considerations as previously discussed, or through more active interventions such as designing alerting systems to promote optimal attention management strategies. Indeed, previous work investigating UAV targeting has shown that active attention cueing through visual, audio, and haptic channels can aid operators in more effectively directing their attention across multiple tasks [89]. Another possible technology solution would be use of some psychophysiological alerting system such as one based on galvanic skin conductance (e.g., [90]) to warn either individual operators or supervisors that a low vigilance state has been achieved. However, as previously discussed, such work is still in early stages. Lastly, how interfaces could be designed to engage and stimulate personnel, such as through on-the-job training or possibly even task-related games, is another area of possible exploration.

**References**

[1]     L. Bainbridge, "Ironies of Automation. Increasing levels of automation can increase, rather than decrease, the problems of supporting the human operator.," *Automatica,* vol. 19, pp. 775-779, 1983.

[2]     J. Langan-Fox, M. J. Sankey, and J. M. Canty, "Keeping the Human in the Loop: From ATCOs to ATMs," in *Keynote Speech by J. Langan-Fox at the Smart Decision Making for Clean Skies Conference*, ed. Canberra, Australia, 2008.

[3]     K. Button, "Different Courses: New Style UAV Trainees Edge Toward Combat," *Training and Simulation Journal,* pp. 42-44, 2009.

[4]     W. T. Thompson, N. Lopez, P. Hickey, C. DaLuz, and J. L. Caldwell, "Effects of Shift Work and Sustained Operations: Operator Performance in Remotely Piloted Aircraft (OP-REPAIR)," presented at the 311th Performance Enhancement Directorate and Performance Enhancement Research Division, Brooks City-Base, Texas, 2006.

[5]     L. J. Prinzel III, H. DeVries, F. G. Freeman, and P. Mikulka, "Examination of Automation-Induced Complacency and Individual Difference Variates," in *NASA/TM-2001-211413*, ed. Langley Research Center: National Aeronautics and Space Administration, 2001.

[6]     J. Langan-Fox, M. J. Sankey, and J. M. Canty, "Human Factors Measurement for Future Air Traffic Control Systems," *Human Factors,* vol. 51, pp. 595-637, 2009.

[7]     T. B. Sheridan, T. Vamos, and S. Aida, "Adapting Automation to Man, Culture and Society," *Automatica,* vol. 19, pp. 605-612, 1983.

[8]     S. Haga, "An experimental study of signal vigilance errors in train driving," *Ergonomics,* vol. 27, pp. 755-765, 1984.

[9]     M. B. Weinger, "Vigilance, Boredom, and Sleepiness," *Journal of Clinical Monitoring and Computing,* vol. 15, pp. 549-552, 1999.

[10]     M. B. Weinger, "Anesthesia Equipment and Human Error," *Journal of Clinical Monitoring and Computing,* vol. 15, pp. 319-323, 1999.

[11]     V. L. Grose, "Coping With Boredom In the Cockpit Before It's Too Late," *Risk Management,* vol. 35, pp. 30-35, 1988.

[12]     J. F. O'Hanlon, "Boredom: Practical consequences and a theory," *Acta Psychologica,* vol. 49, pp. 53-82, 1981.

[13]     (2009, January 12). *Northwest Airlines Flight 188*. Available: http://www.faa.gov/data_research/accident_incident/2009-10-23/

[14]     J. E. See, S. R. Howe, J. S. Warm, and W. N. Dember, "Meta-Analysis of the Sensitivity Decrement in Vigilance," *Psychological Bulletin,* vol. 117, pp. 230-249, 1995.

[15]    S. Straussberger, "Monotony in ATC: Contributing Factors and Mitigation Strategies," Karl-Franzens University, Graz, Austria, 2006.

[16]    S. Straussberger and D. Schaefer, "Monotony in Air Traffic Control," *Air Traffic Control Quarterly, International Journal of Engineering and Operations,* vol. 15, pp. 183-207, 2007.

[17]    C. D. Fisher, "Boredom at work: A neglected concept," *Human Relations,* vol. 46, pp. 395-417, 1993.

[18]    "Random House Webster's Unabridged Dictionary," in *Random House Webster's Unabridged Dictionary*, ed. New York: Random House, 2001.

[19]    P. J. Geiwitz, "Structure of boredom," *Journal of Personality and Social Psychology,* vol. 3, pp. 592-600, 1966.

[20]    V. J. Shackleton, "Boredom and Repetitive Work: A Review," *Personnel Review,* vol. 10, pp. 30-36, 1981.

[21]    R. I. Thackray, "Boredom and Monotony as a Consequence of Automation: A Consideration of the Evidence Relating Boredom and Monotony to Stress," in *DOT/FAA/AM-80/1*, ed. Oklahoma City: Federal Aviation Administration, Civil Aeromedical Institute, 1980.

[22]    R. Farmer and N. D. Sundberg, "Boredom proneness--the development and correlates of a new scale," *Journal of Personality Assessment,* vol. 50, pp. 4-17, 1986.

[23]    D. A. Sawin and M. W. Scerbo, "Effects of Instruction Type and Boredom Proneness in Vigilance: Implications for Boredom and Workload," *Human Factors,* vol. 37, pp. 752-765, 1995.

[24]    A. B. Hill and R. E. Perkins, "Towards a Model of Boredom," *British Journal of Psychology,* vol. 76, pp. 235-240, 1985.

[25]    M. J. Colligan and L. R. Murphy, "Mass psychogenic illness in organizations: An overview," *Journal of occupational psychology,* 1979.

[26]    S. Booth-Kewley, G. E. Larson, R. M. Highfill-McRoy, C. F. Garland, and T. A. Gaskin, "Factors associated with antisocial behavior in combat veterans," *Aggressive Behavior,* vol. 36, pp. 330-337, 2010.

[27]    S. Ebrahim, "Rapid responses, population prevention and being bored to death," *International Journal of Epidemiology,* vol. 39, pp. 323-326, 2010.

[28]    A. Britton and M. J. Shipley, "Bored to death?," *International Journal of Epidemiology,* vol. 39, pp. 370-371, 2010.

[29]    M. L. Cummings. (2008) Of Shadows and white scarves. *C4ISR Journal*.

[30]    D. A. Sawin and M. W. Scerbo, "Vigilance: How to do it and who should do it," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* vol. 2, p. 1312, 1994.

[31]     A. B. Becker, J. S. Warm, W. N. Dember, and P. A. Hancock, "Effects of feedback on perceived workload in vigilance performance," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* p. 1491, 1991.

[32]     M. L. Dittmar, J. S. Warm, W. N. Dember, and D. F. Ricks, "Sex Differences in Vigilance Performance and Perceived Workload," *The Journal of General Psychology,* vol. 120, pp. 309-322, 1993.

[33]     L. J. Prinzel III and F. G. Freeman, "Task-specific sex differences in vigilance performance: subjective workload and boredom," *Perceptual and Motor Skills,* vol. 85, pp. 1195-1202, 1997.

[34]     M. W. Scerbo, C. Q. Greenwald, and D. A. Sawin, "The effects of subject-controlled pacing and task type on sustained attention and subjective workload," *The Journal of General Psychology,* vol. 120, pp. 293-307, 1993.

[35]     A. D. Swain and H. E. Guttman, "Handbook of human reliability analysis with emphasis on nuclear power plant applications," Nuclear Regulatory Commission, Washington D.C. NUREG/CR-1278, 1983.

[36]     G. R. J. Hockey, "Changes in operator efficiency as a function of environmental stress, fatigue and circadian rhythms," in *Handbook of Perception and Human Performance, Vol. II. Cognitive Processes and Performance*, K. R. Boff, L. Kaufmann, and J. E. Thomas, Eds., ed New York: John Wiley and Sons, 1986, pp. 44.1-44.49.

[37]     R. I. Thackray, J. Powell, M. S. Bailey, and R. M. Touchstone, "Physiological, Subjective, and Performance Correlates of Reported Boredom and Monotony While Performing a Simulated Radar Control Task," ed. Oklahoma City: Federal Aviation Administration, Civil Aeromedical Institute, 1975.

[38]     S. J. Kass, S. J. Vodanovich, C. Stanny, and T. M. Taylor, "Watching the Clock: Boredom and Vigilance Performance," *Perceptual Motor Skills,* vol. 92, pp. 969-976, 2001.

[39]     M. D. Rodgers and L. G. Nye, "Factors Associated with Severity of Operational Errors at Air Route Traffic Control Centers," in *An Examination of the Operational Error Database for Air Traffic Control Centers*, M. D. Rodgers, Ed., ed Washington D.C.: Federal Aviation Administration, Office of Aviation Medicine, 1993, pp. 243-256.

[40]     N. H. Mackworth, "Some factors affecting vigilance," *Advancement of Science,* vol. 53, pp. 389-393, 1957.

[41]     E. M. Hitchcock, W. N. Dember, J. S. Warm, B. W. Moroney, and J. E. See, "Effects of Cueing and Knowledge of Results on Workload and Boredom in Sustained Attention," *Human Factors,* vol. 41, pp. 365-372, 1999.

[42]     H. J. Jerison, R. M. Pickett, and H. H. Stenson, "The Elicited Observing Rate and Decision Processes in Vigilance," *Human Factors,* vol. 7, pp. 107-128, 1965.

[43]     M. W. Scerbo, "What's so boring about vigilance?," in *Viewing Psychology as a Whole: The Integrative Science of William N. Dember*, R. R. Hoffman, M. F. Sherrick, and J. S. Warm, Eds., ed: American Psychological Association, 1998, pp. 145-166.

[44]     J. S. Warm, R. Parasuraman, and G. Matthews, "Vigilance Requires Hard Mental Work and is Stressful," *Human Factors,* vol. 50, pp. 433-441, 2008.

[45]     R. Molloy and R. Parasuraman, "Monitoring an Automated System for a Single Failure: Vigilance and Task Complexity Effects," *Human Factors,* vol. 38, pp. 311-322, 1996.

[46]     R. I. Thackray and R. M. Touchstone, "An Evaluation of the Effects of High Visual Taskload on the Separate Behaviors Involved in Complex Monitoring Performance," ed. Oklahoma City: Federal Aviation Administration, Civil Aeromedical Institute, 1988.

[47]     D. J. Schroeder, R. M. Touchstone, N. Stern, N. Stoliarov, and R. I. Thackray, "Maintaining Vigilance on a Simulated ATC Monitoring Task Across Repeated Sessions," ed. Oklahoma City: Federal Aviation Administration, Civil Aeromedical Institute, 1994.

[48]     A. J. Rehmann, "Handbook of Human Performance Measures and Crew Requirements for Flightdeck Research," in *DOT/FAA/CT-TN95/49*, ed. Atlantic City, NJ: Federal Aviation Administration, 1995.

[49]     T. Manly, I. H. Robertson, M. Galloway, and K. Hawkins, "The absent mind: further investigations of sustained attention to response," *Neuropsychologia,* vol. 37, pp. 661-670, 1999.

[50]     N. Pattyn, X. Neyt, D. Henderickx, and E. Soetens, "Psychological investigation of vigilance decrement: Boredom or cognitive fatigue?," *Physiology & Behavior,* vol. 93, pp. 369-378, 2008.

[51]     D. R. Davies and R. Parasuraman, *The Psychology of Vigilance*. London: Academic Press, 1982.

[52]     R. I. Thackray, "The stress of boredom and monotony: a consideration of the evidence," *Psychosomatic Medicine,* vol. 43, pp. 165-176, 1981.

[53]     R. P. Smith, "Boredom: A Review," *Human Factors,* vol. 23, pp. 329-340, 1981.

[54]     T. Cox, "Repetitive work," in *Current concerns in occupational stress*, C. L. Cooper and R. Payne, Eds., ed Chichester, Great Britain: John Wiley & Sons, 1980.

[55]     D. R. Davies, V. J. Shackleton, and R. Parasuraman, "Monotony and boredom," in *Stress and fatigue in human performance*, R. Hockey, Ed., ed Chichester, Great Britain: John Wiley & Sons, 1983, pp. 1-32.

[56]     J. Posner, J. A. Russell, A. Gerber, D. Gorman, T. Colibazzi, S. Yu, Z. Wang, A. Kangarlu, H. Zhu, and B. S. Peterson, "The neurophysiological bases of emotion: An fMRI study of the affective circumplex using emotion-denoting words," *Human Brain Mapping,* vol. 30, pp. 883-895, 2008.

[57]     G. Chanel, C. Rebetez, M. Betrancourt, and T. Pun, "Boredom, engagement and anxiety as indicators for adaption to difficulty in games," in *MindTrek '08 Proceedings of the 12th international conference on Entertainment and media in the ubiquitous era*, New Yourk, NY, 2008.

[58]     H. L. Choi, L. Brunet, and J. P. How, "Consensus-Based Decentralized Auctions for Robust Task Allocation " *IEEE Transactions on Robotics,* vol. 25, pp. 912 - 926, 2009.

[59]     J. J. Gibson, *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin, 1979.

[60]     C. S. Fisher, "Real-Time, Multi-Agent System Graphical User Interface," Master of Engineering, Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 2008.

[61]     M. Valenti, B. Bethke, G. Fiore, J. P. How, and E. Feron, "Indoor Multi-Vehicle Flight Testbed for Fault Detection, Isolation, and Recovery," presented at the AIAA *Guidance, Navigation and Control Conference*, Keystone, CO, 2006.

[62]     W. B. Rouse, *Systems Engineering Models of Human-Machine Interaction*. New York: North Holland, 1983.

[63]     D. K. Schmidt, "A Queuing Analysis of the Air Traffic Controller's Workload," *IEEE Transactions on Systems, Man, and Cybernetics,* vol. 8, pp. 492-498, 1978.

[64]     M. L. Cummings and S. Guerlain, "Developing Operator Capacity Estimates for Supervisory Control of Autonomous Vehicles," *Human Factors,* vol. 49, pp. 1-15, 2007.

[65]     R. W. Proctor and T. V. Zandt, *Human Factors in Simple and Complex Systems*, 2nd ed. Boca Raton, FL: CRC Press, 2008.

[66]     M. L. Cummings and C. E. Nehme, "Modeling the Impact of Workload in Network Centric Supervisory Control Settings," presented at the 2nd Annual Sustaining Performance Under Stress Symposium, College Park, MD, 2009.

[67]     M. S. Sanders and E. J. McCormick, *Human Factors In Engineering and Design*. New York: McGraw-Hill, 1993.

[68]     B. Donmez, C. Nehme, and M. L. Cummings, "Modeling Workload Impact in Multiple Unmanned Vehicle Supervisory Control " *IEEE Systems, Man, and Cybernetics, Part A Systems and Humans,* vol. 99, 2010.

[69]     H. Lomax and N. Casey, "Recording social life: reflexivity and video methodology," *Sociological Research Online,* vol. 3, 1998.

[70]     M. L. Cummings, A. S. Clare, and C. S. Hart, "The Role of Human-Automation Consensus in Multiple Unmanned Vehicle Scheduling," *Human Factors,* vol. 52, 2010.

[71]     M. Jacobs, B. Fransen, J. McCurry, F. W. P. Heckel, A. R. Wagner, and J. Trafton, "A Preliminary System for Recognizing Boredom," in *ACM/IEEE International Conference on Human-Robot Interaction*, La Jolla, CA, 2009, pp. 299-300.

[72]     S. D'Mello, P. Chapman, and A. Graesser, "Posture as a Predictor of Learner's Affective Engagement," in *29th Annual Meeting of the Cognitive Science Society*, Austin, TX, 2007, pp. 905-910.

[73]     J. D. Lee and K. A. See, "Trust in technology: Designing for Appropriate Reliance," *Human Factors,* vol. 46, pp. 50-80, 2004.

[74]     R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A Model for Types and Levels of Human Interaction with Automation," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans,* vol. 30, pp. 286-297, 2000.

[75]     L. D. Loukopoulos, R. K. Dismukes, and I. Barshi, *The Multi-tasking Myth: Handling Complexity in Real-World Operations - (Ashgate Studies in Human Factors for Flight Operations)*. Burlington, VT: Ashgate, 2009.

[76]     E. Ophira, C. Nass, and A. D. Wagner, "Cognitive control in media multitaskers," *Proceedings of the National Academy of Sciences,* vol. 106, 2009.

[77]     J. D. Gould and A. Schaffer, "The effects of divided attention on visual monitoring of multi-channel displays," *Human Factors,* vol. 9, pp. 191-202, 1967.

[78]     D. Tyler and C. Halcomb, "Monitoring performance with a time-shared encoding task," *Perceptual Motor Skills,* vol. 38, pp. 382-386, 1974.

[79]     NYISO, "Interim Report on the August 14, 2003 Blackout," New York Independent System Operator,2004.

[80]     J. Franke, V. Zaychik, T. Spura, and E. Alves, "Inverting the operator/vehicle ratio: Approaches to next generation UAV command and control," presented at the Association for Unmanned Vehicle Systems International and Flight International, Unmanned Systems North America, Baltimore, MD, 2005.

[81]     E. M. Roth, J. Easter, R. E. Hall, L. Kabana, K. Mashio, S. Hanada, T. Clouser, and G. W. Remley, "Person-in-the-Loop Testing of a Digital Power Plant Control Room," *Cognitive Engineering & Decision Making,* pp. 289-293, 2010.

[82]     C. F. Bond and L. J. Titus, "Social facilitation: A meta-analysis of 241 studies," *Psychological Bulletin,* vol. 94, pp. 265-292, 1983.

[83]     D. Guest, R. Williams, and P. Dewe, "Job design and the psychology of boredom," presented at the 19th International Congress of Applied Psychology, Munich, West Germany, 1978.

[84]     J. S. Warm and W. N. Dember, "Awake at the switch," *Psychology Today,* vol. 20, pp. 46-53, 1986.

[85]     C. W. Fontaine and N. D. Schwalm, "Effects of familiarity of music on vigilant performance," *Perceptual and motor skills,* vol. 49, p. 71, 1979.

[86]     P. Bakan, "An analysis of retrospective reports following an auditory vigilance task," in *Vigilance: A symposium*, D. D. Buckner and J. J. McGrath, Eds., ed New York: McGraw Hill, 1963, pp. 88-101.

[87]     R. I. Thackray, K. Jones, and R. M. Touchstone, "Personality and physiological correlates of performance decrement on a monotonous task requiring sustained attention," *British Journal of Social Psychology,* pp. 351-358, 1974.

[88]     S. J. Vodanovich and S. J. Kass, "Age and gender differences in boredom proneness," *Handbook of replication research in the behavioral and social sciences. Special Issue of the Journal of Social Behavior and Personality,* vol. 5, pp. 297-307, 1990.

[89]     D. V. Gunn, J. S. Warm, T. Nelson, R. S. Bolia, D. A. Schumsky, and K. J. Corcoran, "Target Acquisition with UAVs: Vigilance Displays and Advanced Cuing Interfaces," *Human Factors,* vol. 47, pp. 488-497, 2005.

[90]     J. Danckert, "Bored to death: exploring apathetic and agitated subtypes of boredom proneness.," presented at the Cognitive Neuroscience Society, Chicago, IL, 2012.