Adaptive Automation for Multiple Aerial Vehicle Supervisory Control: Impact of Changing Automation Levels

Heath Ruff
Gloria Calhoun
Crystal Miller
Chad Breeden

Follow this and additional works at: https://corescholar.libraries.wright.edu/isap_2013

Part of the Other Psychiatry and Psychology Commons

Repository Citation

This Article is brought to you for free and open access by the International Symposium on Aviation Psychology at CORE Scholar. It has been accepted for inclusion in International Symposium on Aviation Psychology - 2013 by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.
Adaptive automation may help balance system autonomy with human interaction in supervisory control environments. Recent results have demonstrated a benefit of performance-based adaptive automation in a multiple unmanned aerial vehicle simulation. However, these findings may instead reflect an attentional benefit from having the task autonomy level change. A between-subjects experimental design was employed to test this possibility. In one group task autonomy level changed with task performance; in the other group levels changed as a function of time elapsed. The results indicated that performance did not significantly differ between the two groups. However, there were significantly more autonomy level changes in the performance-based adaptive automation group. A follow-on study utilizing a yoked-subject design is recommended.

Advances in automation technology are leading to development of operational concepts in which a single pilot is responsible for multiple unmanned aerial vehicles systems (UASs). In this new vision, the operator’s role will be more supervisory, monitoring the highly automated flight of UASs and updating plans/resource allocations as required in response to changing conditions (Scott, Mercier, Cummings, & Wang, 2006). The increased role of automation, however, can have negative impacts such as reduced operator situation awareness, decision biases, and complacency (Endsley & Kaber, 1999; Sheridan & Parasuraman, 2006). The application of adaptive automation where the system flexibly allocates tasks between the operator and the automation may be useful in optimizing the tradeoff of operator involvement and workload. With adaptive automation, when some triggering condition is met (e.g., based on critical event, operator performance, operator physiology, models of operator cognition, or hybrid method), either higher levels of automation (LOAs) are applied to one or more tasks or the number of tasks automated increases (Prinzel, 2003). For example, with a performance-based adaptive automation scheme, as operator task performance degrades under increased workload/cognitive demands, the autonomy level increases as an aid. As performance improves during the trial, the autonomy decreases to a lower level. This keeps the participant more in the loop and minimizes automation induced problems like complacency.

Past research has demonstrated a benefit of performance-based adaptive automation (e.g., Kaber & Riley, 1999; Parasuraman, Cosenzo, & de Visser, 2009). In more recent research, the utility of a LOA adaptation scheme based on each participant’s individual performance on multiple task types was demonstrated in a multi-UAS simulation (Calhoun, Ward, & Ruff, 2011b; Calhoun, Ruff, Spriggs, & Murray, 2012). Both the speed and accuracy in completing image analysis tasks were improved when this task’s LOA changed across three intermediate levels in response to changes in the participant’s performance. This result was in comparison to data recorded during trials in which the LOA remained constant at the lowest level of automation (Calhoun, et al., 2012). This improved task performance with performance-based adaptive automation may, however, also reflect an attentional benefit from having the autonomy level change during trials. With each change, feedback on LOA state was updated when this task’s LOA changed across three intermediate levels in response to changes in the participant’s performance. This result was in comparison to data recorded during trials in which the LOA remained constant at the lowest level of automation (Calhoun, et al., 2012). This improved task performance with performance-based adaptive automation may, however, also reflect an attentional benefit from having the autonomy level change during trials. With each change, feedback on LOA state was updated in the rightmost console window and a status bar below the map. Also, the configuration of the image analysis task window changed to coincide with the LOA. These changes in the information/appearance of the displays may have had an arousal effect and could have been responsible for past reported improvements in the performance-based adaptive automation condition (Calhoun, et al., 2011b & 2012). The present experiment explored this possibility.

Method

Experimental Design

A mixed-design was utilized. The between-subjects variable was the type of adaptive automation. For one subject group, the autonomy level of an image analysis task was tied directly to the participant’s individual
performance on multiple tasks. In the second subject group, changes to the task’s autonomy level were determined by time into the experimental trial (i.e., not related to task performance). All participants completed three experimental trials with their assigned adaptive-automation condition, as well as three trials with a static automation (image analysis task LOA remained constant during the trial). Each participant’s trials were blocked by automation condition (adaptive and static) and the order of the two trial blocks was counterbalanced across participants.

Twenty-four volunteers served as participants (14 males, 10 females; mean age = 28.79 years, SD = 8.73 years). All participants reported having normal hearing, color vision, and vision (or correctable) to 20/20. None were experienced pilots: 10 were employed at a U.S. Air Force Base and 14 were recruited from a paid subject pool (compensated $15/hr). The participants were randomly and evenly assigned to one of the two automation groups.

Multi-UAV Simulation Apparatus

A testbed developed by OR Concepts Applied was employed as it facilitates experimental manipulation of task LOA (ORCA; Johnson, Leen, & Goldberg, 2007). Also, this Adaptive Levels of Automation (ALOA, Version 3.0) testbed incorporates the ORCA commercially available mission planner to provide needed complexity and realism. The simulation’s computer was a Dell Precision T7500 Workstation with dual Intel® Xeon® CPU x5550 processors @ 2.67 GHz each, 12.0 GB RAM, and a 1.5 GB PCIe nVidia Quadro FX 4800 graphics card (Microsoft© Windows 7 Ultimate 64-bit Operating System). Two Dell 24 in widescreen monitors provided numerous windows that supported participants’ completion of multiple tasks. A keyboard and mouse were used for participant inputs.

Experimental Tasks

Each trial consisted of a series of tasks designed to represent the workload envisioned for multiple autonomous vehicle control. The tasks were also designed such that only a few hours of training were required for naïve participants. More details on each task type and frequency (as well as task order) are available (see Calhoun, Ruff, Draper, & Wright, 2011a). Figure 1 provides an illustration of the formats with labels showing the primary windows utilized for each task. There were approximately 6-7 tasks every minute during each 15-min trial. Some of the task types (change detection, system status, and information retrieval) required monitoring displays and making inputs in response to information displayed. Three other tasks (allocation of imaging tasks to UASs, re-routing UASs, and an image analysis task) employed intermediate LOAs that involved both the operator and automation for completion. For these tasks, the automation was 80% reliable. The LOAs for the allocation and re-routing tasks were constant across trials. In contrast, the image analysis task LOA depended on the automation condition in effect.

Image Analysis Task LOAs and Automation Conditions

Participants were prompted that an image was waiting to be analyzed by the addition of a row in the image task window that included an identifier, time added, vehicle source, and counter showing analysis time remaining. Symbology in a timeline also provided cues of pending images. Participants had 20 s to complete the analysis before the image blanked and the task was recorded as a ‘miss.’ Task completion began with row selection that called up a photo with 19-26 overlaid green shapes (diamonds, squares, circles, and triangles). Analysis required determining the number of diamonds. The next steps depended on the automation condition in effect, to be described next.

Figure 1. Multiple aerial vehicle supervisory control ALOA testbed showing windows used for tasks.
**Static Automation.** In this condition, the LOA for the image analysis task was constant throughout trials. A low LOA was employed in which the automation presented eight options below the image, each with a different number (see Figure 2). Participants were tasked with selecting the option that corresponded to the number of diamonds in the image (1, 2, …or 8). To complete the task and clear the photo, participants clicked “Select.”

**Performance-based Adaptive Automation.** Trials started out with the low LOA. Each time a criterion task type was completed (allocation of image tasks to UASs, routing of UASs, image analysis, and change detection), its corresponding task completion time was compared to an “expected time window.” Instances where task performance was not within thresholds were tallied and once a criterion frequency was met, the LOA changed. (Details on the thresholds are provided in Calhoun, et al., 2012.) As long as the participant’s performance remained within threshold, the LOA stayed at the low level. However, if performance exceeded the thresholds, indicating the participant was over-loaded, the LOA increased to the medium LOA. At this higher level the automation highlighted its recommended option to assist image analysis and reduce cognitive workload. All options were selectable. If the participant agreed with the automation’s recommendation, only the “Select” button needed clicking. However, a different option could be selected. If the participant’s performance indicated that the participant was over-loaded with the medium LOA (i.e., criterion thresholds exceeded), the LOA increased to high. At this level, the automation presented only its recommended option; there was no opportunity to change the selection. To clear the image, the participant’s single requirement was to select “Accept” or “Reject.” If the image timed-out after 20 s, the system recorded an automatic “Accept.” The responses provided a measure of the participant’s detection of automation errors, an indication of complacency.

Increasing LOA in response to performance decline was one part of the adaptive cycle. The LOAs also decreased to re-engage the participant if performance indicated that workload had returned to being manageable. Participants were briefed during training that “the system tracks your performance on several tasks to determine if you are over-loaded or under-loaded and raises or lowers the LOA of the image analysis task in response.”

**Time-based Adaptive Automation.** Trials with time-based adaptive automation also started out with the low LOA. Throughout the trials, the LOA changed, but the changes were unrelated to the participants’ performance. Instead, the changes occurred at experimenter specified times during the trial, and at a similar frequency as that recorded in a previous experiment employing a similar methodology (LOA averaged 3.3 increases and 2.3 decreases across trials; Calhoun, et al., 2012). LOA changes were logical such that, if the LOA was low, a change would make the LOA medium, but it would never immediately transition to high. If the LOA was medium, the LOA might either decrease to low, or increase to high. Participants were briefed on the LOAs and told that the “LOA would change throughout the trial.”

**Procedures**

At the start of the session, each participant completed a background questionnaire and several instruments to measure individual differences (e.g., the 40 Mini-Marker Personality Index to assess the broad traits of the Five Factor Model; Saucier, 2002). A familiarization period followed that took approximately 120 minutes to complete. The testbed’s displays and controls were explained, as well as the scenarios and that the vehicles flew automatically along their flight paths. The automation was described as “reliable, but not perfect.” Next, each task type was described and practiced in the order of the task’s specified priority in a single task environment using the automation condition assigned for the first trial block. This was followed by a series of training trials, gradually increasing the number of task types included in each trial. Training continued until task completion accuracy and response times reached asymptote. Asymptote was defined by mean accuracy and time measures differing by less than 10% on two successive trials that matched the task loading and automation reliability of experimental trials. Next, three 15-min experimental trials with the assigned automation condition were conducted. After each trial, participants completed experimenter developed Likert-type rating scales addressing task difficulty, trust in automation, perceived task performance, situation awareness, workload level, adequacy of automation feedback, and impact of automation on performance. Similar procedures were used for the automation condition assigned for the second trial block. A post-experiment questionnaire was also administered with similar rating scales, as well as questions addressing…
participants’ task completion strategy and comparison of the two automation conditions. The entire session time, including training and questionnaire completion, was approximately 4 hr per participant.

Results

Task Performance Data

The results of an Analysis of Variance (ANOVA) indicated there was not a significant performance difference between the Static (LOA constant at low level) and Adaptive (LOA changed, either performance- or time-based) conditions in terms of task completion time ($F(1,22) = 0.098, p = .757$). Results were similar for image task accuracy, but the means showed a larger difference (static: 70.3%, adaptive: 74.2%; $F(1,22) = 3.692, p = .068$). There were no other significant main effects or interactions. This included task completion time for the participant group that employed the static and performance-based conditions employed in earlier studies ($p > .731$).

ANOVAs were also conducted for other tasks: allocation of image tasks to UASs, routing UASs, and change detection. Mean time to complete the router tasks was faster with the static condition compared to the (performance- and time-based) adaptive condition ($F(1,22) = 4.326, p = .049$). In contrast, mean time to complete the health and status task was longer with the static condition (10.1 s) compared to the adaptive condition (9.4 s; $F(1,22) = 4.607, p = .043$). In a comparison of the two participant groups, mean router task accuracy was better with the performance-based condition (93.9%) compared to the time-based one (88.3%; $F(1,22) = 6.707, p = .017$).

For the performance-based (PB) adaptive automation condition, the mean number of LOA changes for the 12 participants was 8.0 (mean increases and decreases were 4.7 and 3.3, respectively). These data were significantly higher than that employed in the time-based (TB) adaptive condition. Specifically, this was true for the mean total number of changes (8 PB, 5.7 TB; $F(1,22) = 15.66, p = .001$), number of LOA increases (4.7 PB, 3.3 TB; $F(1,22) = 25.919, p < .001$), and number of LOA decreases (3.3 PB, 2.3 TB; $F(1,22) = 8.366, p = .008$). There was also a statistically significant difference between the performance- and time-based adaptive automation conditions in terms of the mean time spent in each of the three LOAs for the image analysis task (Figure 3; $F(2,44) = 7.071, p = .01$). Note that the time spent at each level for the time-based condition was based on experimenter specified adaptive changes. Only the Performance-based Adaptive data were influenced by the participants’ performance.

![Figure 3. Mean time spent in each level of automation during the image analysis task for the performance-based and time-based adaptive automation conditions. Error bars are standard error of the means in this figure and Figure 4.](image)

Subjective Data

Responses significantly differed for only a few ratings. For a post-trial question on how having the image analysis LOA change affected mental workload, an ANOVA of this between-subjects factor indicated that workload was higher with the Performance-based Adaptive Automation compared to Time-based Adaptive ($F(1,22) = 5.782, p = .025$). Responses on a final questionnaire scale comparing Static Automation with the Time-based Adaptive
Automation also differed with 10 of 12 participants indicating that their mental workload was lower with the Time-based Adaptive condition (Kolmogorov-Smirnov non-parametric test: $D(12) = 0.43, p < .05$). On this same questionnaire, participants indicated a preference for the high LOA compared to the other LOAs ($D(12) = 0.55, p < .01$) and rated the frequency in which the LOA changed to be “Sufficient/About Right” (11 of 12 participants; $D(12) = 0.4, p < .05$). This last result is in contrast to a range of responses from the Performance-based Adaptive Automation group: 3 “Slightly Insufficient”, 8 “About Right”, and 1 “Slightly Excessive” ($p > .10$).

**Individual Difference Data**

Individual difference measures were subjected to a median-split procedure in which each participant was classified as either a high (above the median) or low (below the median) responder. Each measure was next analyzed with the median-split category (low versus high) as a between-subject factor, along with the automation conditions described earlier. There were two statistically significant results. One pertained to the Emotion factor (emotional stability; neuroticism; $F(1,20) = 6.032, p = .023$). Post-hoc tests indicated that participants with high Emotion performed the image analysis task slower with the Performance-based Adaptive Automation compared to the Time-based Automation ($t(10) = 2.585, p = .027$; Figure 4). The other result was independent of any automation condition: participants with high Openness were more accurate (77.4%) on the image task analysis compared to participants with low Openness (67.0%; $F(1,20) = 8.651, p = .008$). (Openness has been described as a willingness to engage intellectual challenges, and likely to correlate with better performance (Szalma & Taylor, 2011)).

![Figure 4. Mean time to complete the image analysis task as a function of automation condition (performance- and time-based adaptive) and participants’ Emotion.](image)

**Discussion**

In earlier research with this simulation, participants’ performance on an image analysis task was better with a performance-based adaptive automation control scheme, compared to performance when LOA was static and remained at a low automation level. Thus, it was expected that performance would be better in the present experiment’s performance-based adaptive condition, compared to the time-based condition in which the LOA changes were not tied to real-time performance. The results, though, did not support this hypothesis: performance on the image analysis task did not differ significantly between the two adaptive conditions. This finding suggests that any performance benefit from adaptive automation reflects attention benefits from the display updates associated with the LOA changes. However, the additional result showing that the LOA changed significantly more frequently in the performance-based condition compared to the time-based adaptive condition introduces an alternative explanation. It may be that the number of LOA changes also can influence performance. Even if performance-based LOA changes benefit supervisory control in a multi-task environment, an algorithm that is overly sensitive may activate LOA changes at a frequency that, in turn, hampers performance because it is distracting or irritating. Lending support to this supposition is the tendency of participants in the present study with high emotion to perform worse with the condition with the most LOA changes.

Further research is needed to identify the factors contributing to performance benefits of an adaptive LOA scheme. A variety of algorithms should be examined to compare the benefit of LOA changes tied to performance...
and other trigger alternatives. However, as the present study indicates, the experimental design needs to consider the number of LOA changes as a potentially contributing variable. Use of yoked-subject designs should be considered in which each participant is paired with another on a random basis. For the paradigm used in the present experiment, one participant’s frequency of LOA changes (and timing) could feed the scripted changes for the paired participant, without regard to that participant’s performance. Experimentation that systematically manipulates the threshold criteria used in a performance-based algorithm for LOA changes would also be informative. With such a detailed analysis, the number and timing of LOA changes both within and across participants could more easily be examined. Subsequent research should also evaluate performance-based adaptive control that involves more than one task in order to explore what combinations of LOAs across several tasks is best, without imposing mode awareness issues.

References


