Effects of Unreliable Automation and Individual Differences on Supervisory Control of Multiple Ground Robots

Jessie Y.C. Chen
U.S. Army Research Laboratory
12423 Research Parkway
Orlando, FL 32826, USA
1-407-384-5435
jessie.chen@us.army.mil

Michael J. Barnes
U.S. Army Research Laboratory
2520 Healy Ave, Suite 1172
Ft. Huachuca, AZ 85613, USA
1-520-538-4702
michael.j.barnes@us.army.mil

Caitlin Kenny
University of Central Florida
3100 Technology Parkway
Orlando, FL 32826, USA
1-407-384-3968
ckenny@ist.ucf.edu

ABSTRACT

A military multitasking environment was simulated to examine the effects of unreliable automation on the performance of robotics operators. The main task was to manage a team of four ground robots with the assistance of RoboLeader, an intelligent agent capable of coordinating the robots and changing their routes based upon developments in the mission environment. RoboLeader's recommendations were manipulated to be either false-alarm prone or miss prone, with a reliability level of either 60% or 90%. The visual density of the targeting environment was manipulated by the presence or absence of friendly soldiers. Results showed that the type of RoboLeader unreliability (false-alarm vs. miss prone) affected operator's performance of tasks involving visual scanning (target detection, route editing, and situation awareness). There was a consistent effect of visual density for multiple performance measures. Participants with higher spatial ability performed better on the two tasks that required the most visual scanning (target detection and route editing). Participants' attentional control impacted their overall multitasking performance, especially during their execution of the secondary tasks (communication and gauge monitoring).

Categories and Subject Descriptors

Human Factors

General Terms

Performance, Experimentation, Human Factors.

Keywords

Human Robot Interaction, Simulation, Military, Imperfect Automation, Individual Differences, Multitasking.

1. INTRODUCTION

Unmanned vehicles are being utilized more frequently in military operations, and the tasks they are being used for are evolving in complexity. In the future battlefield, Soldiers may be given multiple tasks to perform concurrently, such as navigating a robot while conducting surveillance, maintaining local security and situation awareness (SA), and communicating with fellow team members. In order to maximize human resources, it is desirable to designate a single operator to supervise multiple robots simultaneously. However, past research has shown that human operators are often unable to control multiple robots/agents simultaneously in an effective and efficient manner [1][2]. As the size of the robot team increases, human operators may fail to maintain adequate SA when their attention is constantly switching between the robots. Cognitive resources may also be overwhelmed by the numerous intervention requests from the robots [3][4]. Wang and his colleagues [4] reviewed a number of studies on supervisory control of multiple ground robots for target detection tasks and concluded that in order to be effective, "the Fan-out plateau lies somewhere between 4 and 9+ robots depending on the level of robot autonomy and environmental demands" (p. 143).

Research shows that autonomous cooperation between robots can aid the performance of the human operators [3] and enhance the overall human-robot team performance [2]. Human operators' involvement in mixed-initiative teams will still be required for the foreseeable future, especially in situations involving critical decision making. Human operators' decision making may be influenced by "implicit goals" that the robots are not aware of (i.e., are not programmed into the behaviors of the robots) [5] and real-time developments on the battlefield that may require the human operator to change plans for individual robots or the entire robotic team. Effective communication between the human operator and robots then becomes critical in ensuring mission success. Carnegie Mellon University researchers demonstrated the effectiveness of a robot proxy to enhance shared understanding between the human operator and the robot in an exploration task [6]. The communication mechanism was based on a common ground collaboration model and improved the human operator performance by assisting in the creation of more accurate plans, more efficient planning (fewer planning repetitions), faster task performance, as well as a better mental model of the capabilities of the robot [6].

1.1 RoboLeader

To achieve a better balance of enhancing autonomy and capability while simplifying human-robot interaction, a robotic surrogate for the human operator, RoboLeader, was developed [7]. RoboLeader is an agent that interprets an operator's intent and issues detailed command signals to a team of robots of lower capabilities. Instead of directly managing the robot team, the human operator deals

HRI'11, March 6–9, 2011, Lausanne, Switzerland.
Copyright 2011 ACM 978-1-4503-0561-7/11/03...$10.00.

with RoboLeader, hypothetically resulting in a reduction of the operator's mental workload. The operator can then better focus on other tasks requiring their attention.

In Chen, Barnes, and Qu [7], the effectiveness of RoboLeader was investigated in a human-in-the-loop simulation experiment. The results show that although there were no significant differences between the RoboLeader and Baseline (no RoboLeader) conditions for target detection performance, participants in the RoboLeader group reduced their mission completion times by approximately 13% compared to the Baseline group. The participants detected significantly fewer targets in the 8-robot condition compared to the 4-robot condition, though participants with higher spatial ability (SpA) detected more targets than did those with lower SpA. Those with lower SpA did not seem to benefit from RoboLeader as much as their higher SpA counterparts. It is likely that the lower SpA participants' scanning of the streaming videos was disrupted by their interaction with RoboLeader whereas the higher SpA participants' scanning was more effective and less affected by their interaction with RoboLeader. When there were 8 robots, participants' SA was significantly worse than when there were only 4 robots. The SA of the RoboLeader participants was not significantly degraded compared with the Baseline group. The "out-of-the-loop" phenomenon associated with automation as reported in previous research [8] did not manifest in the RoboLeader condition. Participants experienced significantly higher workload in the 8-robot condition compared to the 4-robot condition, and those with better attentional control reported lower workload than did those with poorer attentional control.

1.2 Current Study

In the current study, the effects of various reliability levels for RoboLeader on operator performance were investigated. The reliability of RoboLeader's recommendations was manipulated to be either false-alarm prone (FAP) or miss prone (MP), with a reliability level of either 60% or 90%. The effects of imperfect automation are examined by Meyer [9], who suggests that FAP and MP alerts may affect the use of an automated system. High false alarm (FA) rates were seen to reduce the operator's response to alerts (i.e., compliance) while high miss rates reduced the operator's reliance on automated systems. Similar results were reported in Wickens, Dixon, Goh, and Hammer [10]. They found that the operator's *automated* task performance degraded when the automation was FAP due to the operator's reduced compliance with the automated system; when the miss rate was high, the *concurrent* task performance was affected more than the automated task due to the operator allocating more visual attention to monitor the automated task.

In contrast, Dixon, Wickens and McCarley [11] showed that FAP automation hurt "performance more on the automated task than did miss-prone automation, (e.g., the "cry wolf" effect) and hurt performance (both speed and accuracy) at least as much as miss-prone automation on the concurrent task," (p. 11). Similarly, Wickens, Dixon and Johnson [12] demonstrated a greater cost associated with FAP automation (than with MP automation) that affected both the automated and concurrent tasks. Finally, Wickens and Dixon [13] demonstrated that when the reliability level is below approximately 70%, operators will often ignore the alerts. In their meta-analytic study, Wickens and Dixon found that "a reliability of 0.70 was the 'crossover point' below which unreliable automation was worse than no automation at all."

In this study, the effects of individual differences in SpA and attentional control on the operators' robotics control and multitasking performance were evaluated. Lathan and Tracey [14] found that people with higher SpA performed faster and with fewer errors in a robot teleoperation task. It was suggested that military missions can benefit from selecting personnel with higher SpA to operate robotic devices. Previous studies have also found SpA to be a good predictor of the operator's robotics performance [1][7]. In the previous RoboLeader study [7], participants with higher SpA scanned the videos significantly faster than those with lower SpA.

The relationship between perceived attentional control (PAC) and multitasking performance was also evaluated in this study. Research has shown that there are individual differences in multitasking performance; some people are less prone to performance degradation during multitasking conditions [15]. There is evidence that people with better attentional control can allocate their attention more flexibly and effectively [16][17]. A strong interaction between the type of automation unreliability and participants' attentional control was found by Chen [18], which investigated operator's multitasking performance of gunnery and robotics tasks. Overall, for high PAC participants, FAP alerts were more detrimental than MP alerts. High PAC participants tended to rely on their own multitasking ability instead of relying on the automated systems. For low PAC participants, MP automation was more harmful than FAP automation. Low PAC participants relied on the MP automation more than they should have; their trust in the FAP system resulted in better performance than the high PAC participants'. The current study sought to examine whether individuals with different attentional control abilities interacted differently with FAP and MP RoboLeader.

2. METHOD

2.1 Participants

Forty individuals (23 males and 17 females, mean age 23.8 yrs) from the Orlando, FL area participated in the study. They were compensated $15/hr for their time.

2.2 Apparatus

2.2.1 Simulator

A modified version of the Mixed Initiative Experimental (MIX) Testbed was used as the simulator for this experiment. The MIX Testbed is a distributed simulation environment for investigation into how robots are used and how automation affects human operator performance [19]. The Operator Control Unit (OCU) of the MIX Testbed (Fig. 1) was modeled after the Tactical Control Unit developed under the ARL Robotics Collaborative Technology Alliance (CTA). This platform includes a camera payload and supports multiple levels of automation. Users can send mission plans or teleoperate the platform with a computer mouse while being provided video feeds from the camera payload. Typical tasks include reconnaissance and surveillance. RoboLeader has the capability of collecting information from subordinate robots with limited autonomy (e.g., collision avoidance and self-guidance to reach target locations), making tactical decisions and coordinating the robots by issuing commands, waypoints, or motion trajectories [7][20].

Figure 1. RoboLeader user interface.

2.2.2 *Simulation of Unreliable RoboLeader*

In Chen et al. [7], the simulated reliability level of RoboLeader was 100% (i.e., no FAs or misses). In the current study, the reliability of RoboLeader was either FAP or MP, at a level of either 60% or 90%. This created four different RoboLeader reliability conditions: FAP60, FAP90, MP60, and MP90 (see Procedure).

2.3 Surveys and Tests

Before the training session, a series of surveys and tests were given to the participants, beginning with a demographic questionnaire. An Ishihara Color Vision Test with 9 test plates was administered via PowerPoint presentation to ensure that the participants' color vision was normal. The Attentional Control Survey [16] was used to evaluate participants' perceived attentional control. The Attentional Control Survey consists of 20 items and measures attention focus and shifts. The scale has been shown to have good internal reliability (α = .88). The Cube Comparison Test [21] was used to assess participants' spatial ability. The Cube Comparison Test requires participants to compare, in 3 min, 21 pairs of 6-sided cubes and determine if the rotated cubes are the same or different.

Participants' perceived workload was evaluated using the computer-based version of NASA-TLX questionnaire [22] after each experimental condition. A modified version of the Usability and Trust Questionnaire used in Chen [18] assessed participants' perceived usability of the RoboLeader system as well their trust in the system. The items that measured participants' trust in the system were modified from the "Trust Between People and Automation" questionnaire [23]. The questionnaire consists of 22 questions on a scale of 1 to 7 and includes questions such as "The RoboLeader display can be deceptive," and "The RoboLeader system is dependable."

2.4 Experimental Design

The study is a mixed design, with Unreliability Type (FAP vs. MP) and Reliability Level (60% [Low] vs. 90% [High]) as the between-subject factors, and Visual Density of scenario (High [with friendly soldiers in the scenarios] vs. Low [without friendly soldiers in the scenarios]) as the within-subject variable (see Procedure).

2.5 Procedure

Participants were randomly assigned to the FAP60, FAP90, MP60, or MP90 group (with 10 participants per group) before their sessions started. After being briefed on the purpose of the study and signing the informed consent form, participants completed the Demographics Questionnaire, the Attentional Control Survey, the Ishihara Color Vision Test and the Cube Comparison Test. Participants then received training and practice on the tasks they were about to conduct during the experimental session. Training was self-paced and delivered by PowerPoint® slides showing the elements of the OCU, steps for completing various tasks, several mini-exercises for practicing the steps and exercises for performing the robotic control tasks. The type and reliability level of each participant's RoboLeader condition (FAP60, FAP90, MP60, or MP90) matched in the training and experimental scenarios. The participants were informed that RoboLeader was either FAP or MP and "fairly but not always reliable" (for the 90% conditions) or "not always reliable" (for the 60% conditions). Before proceeding to the experimental session, participants had to demonstrate that they could recall all the steps for performing the tasks without any help. The training session lasted approximately 1 hour.

The experimental session began immediately after the training session and lasted about 1 hour. Each experimental session had two scenarios, both lasting approximately 30 minutes. The order of scenarios was counterbalanced across participants. During the scenarios, participants used their 4 robots to locate 10 targets (insurgents carrying weapons) while rerouting their robots around events in the remote environment (described later). When each scenario started, the robots began to follow pre-planned routes and operators began to monitor the environment for target (by clicking on one of the four video thumbnail views highlighted in the color of its route to enlarge it into the expanded view on the top the screen) and event detection. The robots did not have Aided Target Recognition capability and the participants had to detect the 10 insurgents by themselves. To identify targets, participants used a mouse to click the Insurgent button on the interface and then click directly on the insurgent to "laze" them as soon as they were detected. The "lazed" insurgent was then displayed on the map by a white, diamond shaped "THREAT" icon (see Fig. 1). In the low density scenario, there were about 600 civilians throughout the scenario and in the high density scenario there were about 600 civilians and 600 friendly soldiers visible in the environment. The presence of friendly soldiers in the high density scenario made the target detection task more difficult as the friendly soldiers all carry weapons.

During the scenarios, several "events" (notifications that the human operator receives from the intelligence network stating that certain areas were High Priority or Hostile) required revisions to the robots' routes. RoboLeader and the participants needed to create new routes towards High Priority Areas while avoiding re-routing the robots through Hostile Areas or areas already traversed. Once an event transpired (indicated by appearance of an icon on the map), the participants needed to notice and acknowledge that the event had occurred. RoboLeader then recommended route revisions for the events (by presenting the new route to the operator visually on the map) which the operator either accepted, or rejected and modified as deemed necessary.

In the MP scenarios, participants were required to notice and manually edit several routes without the help of RoboLeader. RoboLeader's messages were displayed in the blue, upper left corner of the OCU (see Fig. 1). Participants were told that their objective was to finish reconnoitering the area using their robots in the least amount of time possible while keeping all route edits as close as possible to the original routes. When re-planning a route, the participants and RoboLeader were required to consider the effectiveness and efficiency of a new route. Situations where a robot completes its route rapidly but does not cover much ground, or when the robot covers a lot of ground but is slow to finish would be suboptimal.

In the FAP60 scenario, following the Signal Detection Theory paradigm [12][24], there were five true events that required revisions to a robot's route and four FAs that RoboLeader attempted to edit around when no events occurred (zero misses and one correct rejection), making a total of ten events, six of which were positive. Participants could check the validity of RoboLeader's recommendations by reviewing their map. A true event was associated with an icon (a red square for a Hostile Area and a blue square for a High Priority Area, see Fig. 1), but FAs were not. In the FAP90 scenario, there were five true events that required revisions to a robot's route, and one FA. In the MP60 scenario, ten true events occurred that required revisions to a robot's route, though RoboLeader only provided solutions for two of them. In the MP90 scenario, ten true events occurred and RoboLeader provided solutions for eight of them.

Each scenario also contained five situation awareness (SA) queries, which were triggered based on time progression (e.g., 3 min into the scenario). The SA queries included prompts such as "Identify which routes have encountered the most Insurgents." When an SA query was triggered, the OCU screen went blank with the simulation paused and the SA query displayed on the screen. Participants then wrote their response on an answer sheet. After finished writing, participants clicked a "Resume" button on the screen to continue the experiment.

In addition to the robotics tasks described above, the participants simultaneously performed a gauge monitoring task and an auditory communications task. The gauge monitoring task (upper left corner of the OCU under the blue RoboLeader message box) displayed four gauges constantly in motion that entered an upper or lower limit at various pre-specified times throughout the scenarios. The participants were required to monitor the gauges and press a "Reset," button when any gauge entered the upper or lower limit to put the gauges back to their normal levels. The auditory communications task presented pre-recorded questions at 30 second intervals during the scenarios. Questions included simple military-related reasoning and memory tests. For the reasoning tests, questions such as, "If the enemy is to our left and our robot is to our right, what direction is the enemy to the robot?" were asked. For the memory tests, participants performed a recognition task for three radio call signs (Alpha 27, Bravo 45 and Bravo 83) and determined whether each call sign they heard was one of the three specified. Participants used a keyboard to enter their responses for the questions into the communications panel on the OCU (adjacent to the gauges, see Fig. 1).

A 2-minute break was given between the experimental scenarios. Participants assessed their perceived workload using an electronic NASA-TLX immediately after each scenario. Following completion of both scenarios, participants were asked to evaluate the usability of the RoboLeader system by filling out the Usability and Trust Questionnaire.

2.6 Measures and Data Analysis

Dependent measures include the number of targets located and identified, the number of routes successfully edited, the operators' SA of the mission environment, concurrent task performance (gauge monitoring and auditory communications) and perceived workload. A mixed design ANCOVA with Unreliability Type (FAP vs. MP) and Reliability Level (60% [Low] vs. 90% [High]) as the between-subject factors and Visual Density (High vs. Low) as the within-subject factor is used to evaluate the operators' performance differences among the four conditions. Participants' spatial ability (Cube Comparison Test score) and their perceived attentional control (PAC; Attentional Control Survey score) are used as covariates.

3. RESULTS

3.1 Operator Performance

A multivariate ANCOVA was first performed to assess the overall effects of the factors on all five performance measures (target detection, routes edited, SA, communication, gauge monitoring). The analysis showed that there was a main effect of Unreliability Type, $F(5,30) = 59.0$, $p < 0.0001$, $\eta^2_p = 0.91$. There was also a significant difference between those with higher and lower PAC, $F(5,30) = 3.4$, $p < 0.05$, $\eta^2_p = 0.36$. There was a significant 3-way interaction among Unreliability Type, Reliability Level, and Visual Density, $F(5,30) = 2.8$, $p < 0.05$, $\eta^2_p = 0.32$. The effect of Reliability Level, Visual Density, and operator spatial ability (SpA) failed to reach statistical significances, p's > 0.05. Analysis on each performance measure is summarized below.

3.1.1 Target Detection Performance

The analysis of the operators' target detection performance revealed that there was a main effect of Unreliability Type and Visual Density, $F(1,35) = 45.7$, $p < 0.0001$, $\eta^2_p = 0.57$ and $F(1,35) = 12.2$, $p < 0.001$, $\eta^2_p = 0.26$, respectively. Participants detected significantly fewer insurgents in the MP condition than in the FAP condition and detected significantly fewer insurgents in the High Density environment than in the Low Density environment (Fig. 2).

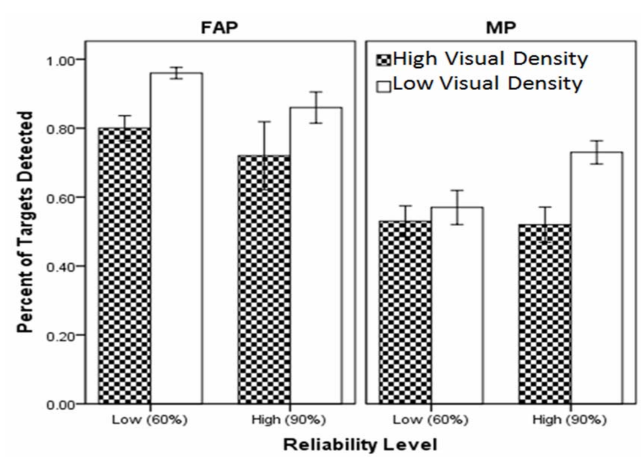

Figure 2. Target detection.

There was a significant difference between those with a higher SpA ability and those with a lower SpA ability, $F(1,35) = 4.1$, $p < 0.05$, $\eta^2_p = 0.104$ (Fig. 3).

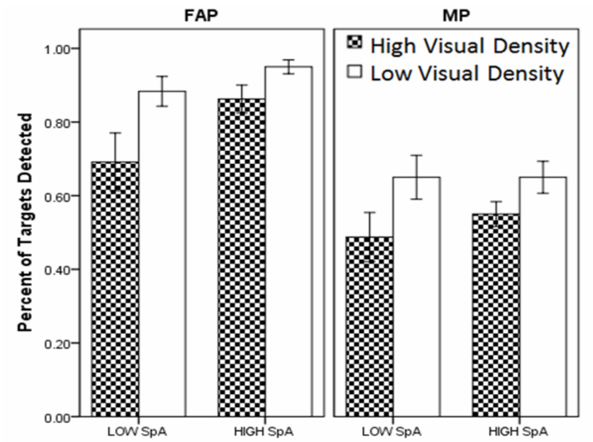

Figure 3. Effects of SpA and visual density on target detection.

3.1.2 Route Editing

The analysis showed that Unreliability Type and Reliability Level of RoboLeader significantly affected the percentage of routes that participants successfully edited, $F(1,35) = 161.7$, $p < 0.0001$, $\eta^2_p = 0.82$ and $F(1,35) = 7.4$, $p < 0.01$, $\eta^2_p = 0.18$, respectively (Fig. 4). Participants edited more routes in the FAP condition than in the MP condition and successfully edited more routes in the High Reliability (90%) condition than in the Low Reliability (60%) condition (Fig. 4). There was a significant difference between those with a higher SpA ability and a lower SpA ability, $F(1,35) = 7.5$, $p < 0.01$, $\eta^2_p = 0.18$.

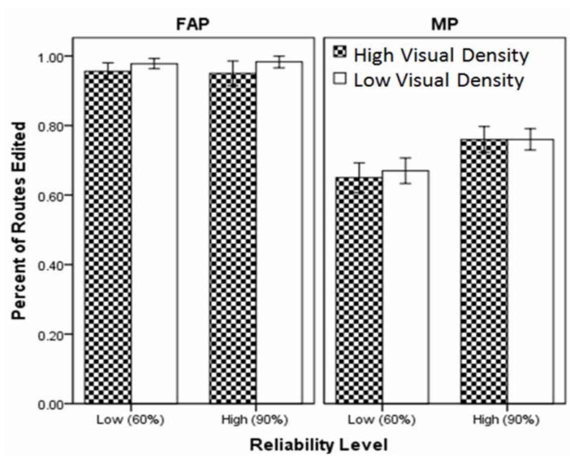

Figure 4. Routes edited.

3.1.3 Situation Awareness

The Unreliability Type of RoboLeader significantly affected the participants' SA of the mission environment (i.e., number of SA queries answered correctly), $F(1,35) = 8.5$, $p < 0.01$, $\eta^2_p = 0.20$ (Fig. 5). Participants' SA was significantly better in the MP condition than in the FAP condition.

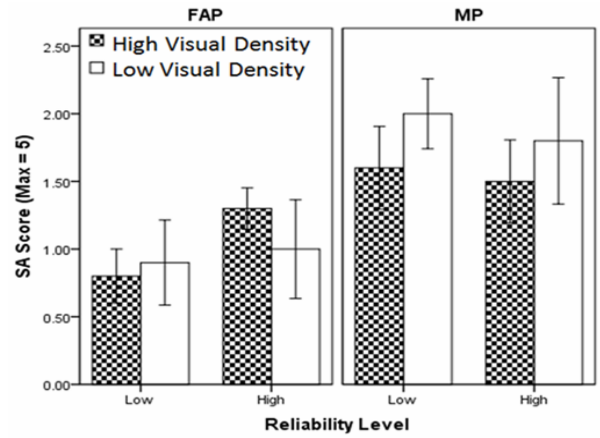

Figure 5. Situation awareness.

3.1.4 Communication Task Performance

Visual Density of the target environment significantly affected participants' communication task performance, $F(1,35) = 5.8$, $p < 0.05$, $\eta^2_p = 0.14$. Participants' communication task performance was slightly better in the High Visual Density condition than in the Low Visual Density condition. There was a significant 3-way interaction among Visual Density, Type of RoboLeader Unreliability, and Reliability Level, $F(1,35) = 10.3$, $p < 0.01$, $\eta^2_p = 0.23$, and a significant interaction between Visual Density and participants' PAC, $F(1,35) = 5.4$, $p < 0.05$, $\eta^2_p = 0.134$.

3.1.5 Gauge Monitoring Performance

There was a significant interaction between Visual Density and Reliability Level of RoboLeader, $F(1,35) = 4.3$, $p < 0.05$, $\eta^2_p = 0.11$. Participants' gauge monitoring task performance (response times) tended to be better (faster) in the Low Visual Density when the Reliability Level of RoboLeader was High; however, when RoboLeader was not reliable, participants tended to respond faster in the High Visual Density condition. There was a significant difference between participants with low PAC and high PAC, $F(1,35) = 84.7$, $p < 0.0001$, $\eta^2_p = 0.31$, with High PAC participants responding faster than those with low PAC (Fig. 6).

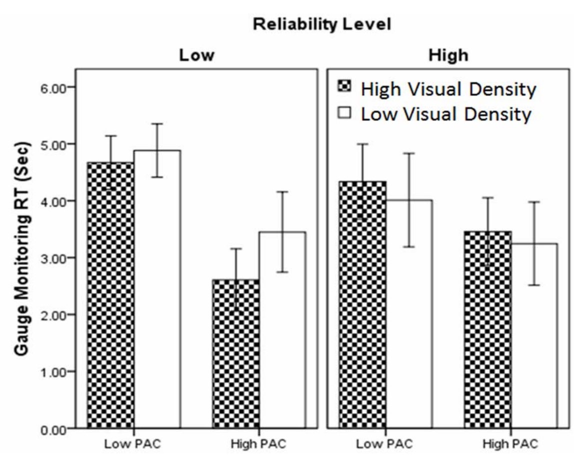

Figure 6. Effects of PAC on gauge monitoring performance.

3.2 Perceived Workload

The analysis showed that both Visual Density of the target environment and the Reliability Level of RoboLeader contributed significantly to the participants' perceived workload, $F(1,36) = 7.7$, $p < 0.01$, $\eta^2_p = 0.18$ and $F(1,36) = 4.8$, $p < 0.05$, $\eta^2_p = 0.12$ respectively (Fig. 7). Participants experienced higher workload in the High Density condition as well as when the Reliability Level was lower.

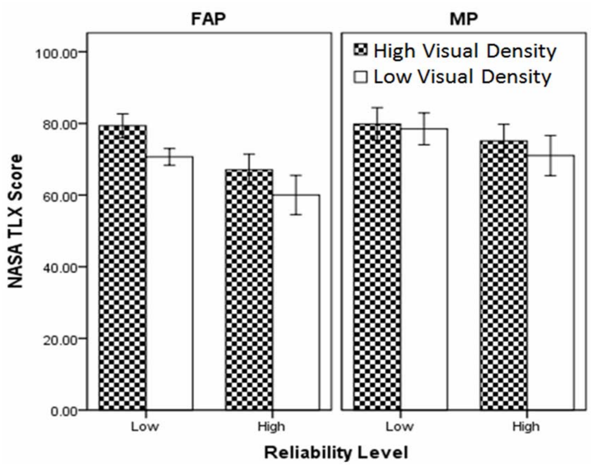

Figure 7. Perceived workload.

3.3 Operators' Interaction with the OCU

Participants' interaction with the OCU (clicks made on the OCU) was analyzed. Participants made significantly more thumbnail clicks in the Low Visual Density condition than in the High Density condition, $F(1,35) = 6.5$, $p < 0.05$, $\eta^2_p = 0.16$. Participants' spatial ability (Cube Comparison Test score) was also significantly correlated with the number of thumbnail clicks, $r = 0.33$, $p < 0.05$. The participants' self-assessed trust in the RoboLeader system was also evaluated and those with higher and lower PAC were compared. There was no significant difference between these two groups in the aggregate scores of the Trust questionnaire, $p > 0.05$. Participants' responses on some items did show differences among the groups. For the items "I can trust the RoboLeader system" and "The RoboLeader display can be deceptive," the FAP group rated them significantly higher than the MP group, p's < 0.05. For the item "I relied heavily on the RoboLeader for the task," participants in the High Reliability group rated it significantly higher than did those in the Low Reliability group, $p < 0.05$.

4. DISCUSSION

Overall, there was a consistent effect of Type of Unreliability (FAP vs. MP) for tasks involving visual scanning (target detection, route editing and SA) with a distinct pattern of tradeoff among the measures. Participants performed more poorly with the MP RoboLeader for the target detection task and the route editing tasks, but maintained better SA with the MP RoboLeader. This suggests that the participants scanned the map more frequently in the MP condition than in the FAP condition. This is consistent with the findings of Wickens, Dixon, Goh, and Hammer [10] that MP systems drew operators' visual attention away from the concurrent tasks to focus more on the automated tasking environment. Participants in the MP group also reported that they trusted the RoboLeader significantly less than those in the FAP group. Interestingly, the FAP participants thought the RoboLeader was more "deceptive" than the MP participants.

The Reliability Level of RoboLeader was found to have an effect on route editing and operators' perceived workload. The first finding was not surprising - participants successfully edited more routes in the High Reliability (90% accuracy) condition than in the Low Reliability (60% accuracy) condition. Participants also experienced higher workload and reported significantly less reliance on the RoboLeader with the Low Reliability condition indicating that participants could discriminate between these two Reliability conditions and act accordingly.

There was a consistent effect of Visual Density for multiple performance measures: target detection, communication, numbers of thumbnail clicks, and perceived workload. As expected, participants made more thumbnail clicks and detected more targets in the Low Density condition. Interestingly, participants performed slightly better on the Communication task in the High Density condition, although the effect was not straightforward and there was a complex interaction among Visual Density, Type of RoboLeader Unreliability, and Reliability Level. Participants' perceived workload was significantly higher when the Visual Density was higher.

Participants with higher spatial ability performed better on the two tasks that required the most visual scanning: target detection and route editing. They also made more thumbnail clicks, consistent with what was found in the previous experiment on RoboLeader [7]. These findings suggest that those with higher spatial ability were able to scan the tasking environment faster than those with lower spatial ability. Endsley and Bolstad [25] found that pilots with higher spatial ability were significantly better able to acquire and maintain SA than their lower spatial ability counterparts. While a similar correlation was not found in the SA measure in the current study, our findings were consistent with Endsley and Bolstad's since target detection and route editing (i.e., change detection) can be considered as Level 1 (perception) and Level 2 (comprehension) SA of the mission environment [26]. The findings also support the recommendations by Lathan and Tracey [14] and two recent U.S. Air Force studies [27][28] that military missions can benefit from selecting personnel with higher spatial ability to operate robotic devices. Training interventions that could enhance the spatial interpretations required to perform a mission task might also be of benefit [29].

Participants' perceived attentional control (PAC) was found to have a significant effect on their multitasking performance, especially the execution of secondary tasks (communication and gauge monitoring). This finding is consistent with those of Chen and Joyner [17] that participants performed at a similar level on the primary tasks (gunnery and robotics), but those with higher PAC performed better on the secondary communication task than did those with lower PAC. These results suggest that participants with higher PAC were more able to allocate their attentional resources in the multitasking environment than those with lower PAC. It was also found that participants with higher PAC consistently performed better in the MP condition across different tasks than those with lower PAC. This is consistent with Chen

[18] that MP automated systems tended to be more detrimental to lower PAC individuals than to higher PAC individuals.

An interesting difference between the current results and those of Chen [18] was that in the current study, participants with higher PAC did not exhibit as much under-trust (i.e., disuse) of the FAP system as those high PAC participants did in the Chen [18] study. In the current study, high PAC participants performed at similar levels as low PAC participants in the FAP conditions, but outperformed low PAC individuals in the MP conditions. The discrepancy between these results and those of Chen [18] may be due to the different "costs" of scanning in the two simulated environments. In the Chen [18] study, the gunner station and the robotics OCU were displayed on two separate monitors, while in the current study all tasks were performed on the same monitor. Compared to the current study, the cost of scanning in the Chen study was greater and those of higher PAC clearly demonstrated reduced compliance with the FAP automated system. In the current study, high PAC participants did not show this decrement, likely due to the relative ease of verifying the RoboLeader recommendations on the map by checking the icons.

It is interesting to note that while it was considerably easier to verify the validity of the alerts in the current study, participants with low PAC performed more poorly in the MP conditions than those with high PAC, just as the results of the Chen [18] study showed. A likely reason for this phenomenon is that MP scenarios required continuous scanning of the map to find new icons. This made the task similar to a "change detection" task, though performed in a multitasking environment. The current results suggest that low PAC individuals cannot detect changes as effectively as their high PAC counterparts likely due to their poorer attentional management abilities. The way the low PAC participants interacted with the automated system in the current experiment was consistent with the "cognitive miser" phenomenon described in Feldman Barrett, Tugade and Engle [30]. The phenomenon states that low PAC individuals, due to their limited attentional resources, tend to reduce their information processing demands by simplifying their task(s) (e.g., relying on RoboLeader to help them with their plan revision tasks). Depending on the context, this over-simplification (i.e., over-reliance on automation) may have very undesirable consequences (e.g., MP condition) when the aids fail to provide anticipated assistance.

5. CONCLUSIONS

In the current study, we investigated the effects of unreliable automation (FAP vs. MP) on human operators' performance of supervising multiple robots to complete military reconnaissance missions and their concurrent performance of communication and gauge monitoring tasks. Results show that the type of RoboLeader unreliability affected operator's performance of tasks involving visual scanning (target detection, route editing, and situation awareness). There was a consistent effect of visual density for multiple performance measures. Participants with higher spatial ability performed better on the two tasks that required the most visual scanning (i.e., target detection and route editing). Participants' self-assessed attentional control was found to impact their overall multitasking performance, especially in execution of the secondary tasks (communication and gauge monitoring). The current study also presents some evidence that spatial proximity between visual displays does exert an effect on both automated and concurrent task performance [31]. The implications of the findings of the current study to the military robotics are manifold. The requirement for many-to-one supervision is a practical means to increase force capability while maintaining acceptable manning levels; robot-to-robot interactions or capabilities similar to RoboLeader can potentially achieve the goal of many-to-one supervision for the human operator. Successful implementation of robot-to-robot control can seed future research for such programs as the U.S. Army Research Laboratory's Robotics CTA as well as supplement the autonomous capabilities being developed for the Safe Operations for Unmanned Reconnaissance in Complex Environments (SOURCE) Army Technology Objective (ATO). However, since no automation can achieve 100% reliability at all times, the effects of unreliable automated systems on human operator performance needs to be better understood before these intelligent systems can be implemented. Finally, the current study shows that individual differences such as spatial ability and attentional control can have a profound impact on operator's task performance. Training programs (e.g., attention management) and/or user interface designs (e.g., multimodal cueing displays) should be developed to mitigate performance shortfalls of those with lower spatial ability and attentional control [32].

6. ACKNOWLEDGMENTS

This research was funded by U.S. Army's SOURCE ATO. RoboLeader was developed under a U.S. Army Research Laboratory (ARL) Director's Research Initiative (DRI) project in FY09. The authors wish to thank Zhihua Qu, Mark Snyder, and Daniel Barber of Univ. of Central Florida for their contributions.

7. REFERENCES

[1] Chen, J. Y. C., Durlach, P., Sloan, J., and Bowens, L. 2008. Human robot interaction in the context of simulated route reconnaissance missions. *Mil. Psychol.* 20, 3 (July 2008), 135-149. DOI=10.1080/08995600802115904.

[2] Schurr, N. 2007. *Toward human-multiagent teams.* Doctoral Thesis. Pub No. 3291875, Univ. of Southern California.

[3] Wang, H., Lewis, M., Velagapudi, P., Scerri, P., and Sycara, K. 2009. How search and its subtasks scale in N robots. In *Proc. 4th ACM/IEEE Int. Conf. on Human-Robot Interaction* (La Jolla, CA, Mar 11-13, 2009). HRI '09. ACM, New York, 141-147. DOI= 10.1145/1514095.1514122.

[4] Wang, J., Wang, H., and Lewis, M. 2008. Assessing cooperation in human control of heterogeneous robots. In *Proc. 3rd ACM/IEEE Int. Conf. on Human-Robot Interaction* (Amsterdam, the Netherlands, Mar. 12-15, 2008). HRI '08. ACM, New York, 9-15. DOI= 10.1145/1349822.1349825.

[5] Linegang, M., Stoner, H., Patterson, M., Seppelt, B., Hoffman, J., Crittendon, Z., and Lee, J. 2006. Human-automation collaboration in dynamic mission planning: A challenge requiring an ecological approach. In *Proc. 50th Human Factors and Ergonomics Society Annual Meeting* (San Francisco, CA, Oct. 16-20, 2006). HFES '06. HFES, Santa Monica, CA, 2482-2486.

[6] Stubbs, K., Wettergreen, D., and Nourbakhsh, I. 2008. Using a robot proxy to create common ground in exploration tasks," In *Proc. 3rd ACM/IEEE Int. Conf. on Human-Robot*

Interaction (Amsterdam, the Netherlands, Mar. 12-15, 2008). HRI '08. ACM, New York, 375-382.

[7] Chen, J. Y. C., Barnes, M. J., and Qu, Z. 2010. RoboLeader: An agent for supervisory control of multiple robots. In *Proc. 5th ACM/IEEE Int. Conf. on Human-Robot Interaction* (Osaka, Japan, Mar 2-5, 2010). HRI '10. ACM, New York, 81-82.

[8] Endsley, M. and Kiris, E. 1995. The out-of-the-loop performance problem and level of control in automation. *Human Factors* 37 (Jun 1995), 381-394.

[9] Meyer, J. 2004. Conceptual issues in the study of dynamic hazard warnings. *Human Factors* 46 (Summer 2004), 196-204.

[10] Wickens, C., Dixon, S., Goh, J., and Hammer, B. 2005. *Pilot dependence on imperfect diagnostic automation in simulated UAV flights: An attentional visual scanning analysis.* Tech Report: AHFD-05-02/MAAD-05-02, Univ. of Illinois, Urbana-Champaign, IL, 2005.

[11] Dixon, S., Wickens, C., and McCarley, J. 2007. On the independence of compliance and reliance: Are automation false alarms worse than misses? *Human Factors* 49 (Aug. 2007), 564-572

[12] Wickens, C., Dixon, S., and Johnson N. 2005. *UAV automation: Influence of task priorities and automation imperfection in a difficult surveillance task*. Tech Report: AHFD-05-20/MAAD-05-6, Univ. of Illinois, Urbana-Champaign, IL, 2005.

[13] Wickens, C., and Dixon, S. 2005. *Is there a magic number 7 (to the minus 1)? The benefits of imperfect diagnostic automation: A synthesis of the literature.* Tech Report: AHFD-05-01/MAAD-05-01, Univ. of Illinois, Urbana-Champaign, IL, 2005.

[14] Lathan, C. and Tracey, M. 2002. The effects of operator spatial perception and sensory feedback on human-robot teleoperation performance. *Presence* 11 (Aug. 2002), 368-377.

[15] Rubinstein, J., Meyer, D., and Evans, J. 2001. Executive control of cognitive processes in task switching. *J. Exp. Psychol.: Human Perception and Performance* 27 (Aug. 2001), 763-797.

[16] Derryberry, D. and Reed, M. 2002. Anxiety-related Attentional Biases and Their Regulation by Attentional Control. *J. Abnormal Psychol.* 111 (May 2002), 225-236.

[17] Chen, J. Y. C. and Joyner, C. 2009. Concurrent performance of gunner's and robotic operator's tasks in a multi-tasking environment. *Mil. Psychol.* 21, 1 (Jan. 2009), 98 – 113.

[18] Chen, J. Y. C. 2009. Concurrent performance of military tasks and robotics tasks: Effects of automation unreliability and individual differences. In *Proc. 4th ACM/IEEE Int. Conf. on Human-Robot Interaction* (La Jolla, CA, Mar 11-13, 2009). HRI '09. ACM, New York, 181-188. DOI=10.1145/1514095.1514128.

[19] Barber, D., Davis, L., Nicholson, D., Finkelstein, N, and Chen, J. Y. C. 2008. The mixed initiative experimental (MIX) testbed for human robot interactions with varied levels of automation. In *Proc. 26th Army Sci. Conf* (Orlando, FL, Dec. 1-4, 2008). ASC '08. US Dept. of Army, Washington, DC.

[20] Snyder, M., Qu, Z., Chen, J. Y. C., and Barnes, M. 2010. RoboLeader for reconnaissance by a team of robotic vehicles. In *Proc. Int. Symp. on Collaborative Technologies and Systems* (Chicago, May 17-21, 2010). CTS '10. IEEE, New York, 522-530.

[21] Ekstrom, R., French J., and Harman, H. 1976. *Kit of Factor-Referenced Cognitive Tests.* Educational Testing Service, Princeton, NJ.

[22] Hart, S., and Staveland, L. 1988. Development of NASA TLX (Task Load Index): Results of empirical and theoretical research. In *Human Mental Workload*, P. Hancock & N. Meshkati, Eds. Elsevier, Amsterdam, 139-183.

[23] Jian, J., Bisantz, A., and Drury, C. 2000. Foundations for an empirically determined scale of trust in automated systems. *Int. J. Cognitive Ergonomics* 4 (Jan. 2000), 53-71.

[24] Green, D. and Swets, J. 1988. *Signal Detection Theory and Psychophysics.* Wiley, New York.

[25] Endsley, M. and Bolstad, C. 1994. Individual differences in pilot situation awareness. *Int. J. Aviation Psychol.* 4 (July 1994), 241-264.

[26] Endsley, M. 1995. Toward a theory of situation awareness in dynamic systems. *Human Factors* 37 (Mar 1995), 32–64.

[27] Chappelle, W., McMillan, K., Novy, P., and McDonald, K. 2010. Psychological profile of USAF unmanned aerial systems Predator and Reaper pilots. *Aviat., Space, & Envir. Med.* 81 (May 2010), 339.

[28] Chappelle, W., Novy, P., Randall, B., and McDonald, K. 2010. Critical psychological attributes of U.S. Air Force (USAF) Predator and Reaper sensor operators according to subject matter experts. *Aviat., Space, & Envir. Med.* 81 (May 2010), 253.

[29] Rodes, W., Brooks, J., and Gugerty, L. 2005. Using verbal protocol analysis and cognitive modeling to understand strategies used for cardinal direction judgments. *Poster presented at the Human Factors of UAVs Workshop* (Mesa, AZ, May 25-26, 2005).

[30] Feldman Barrett, L., Tugade, M., and Engle, R. 2004. Individual differences in working memory capacity and dual-process theories of the mind. *Psychol. Bull.* 130 (July 2004), 553-573.

[31] Wickens, C. and Hollands, J. 2000. *Engineering Psychology and Human Performance* (3rd ed.). Prentice Hall, Upper Saddle River, NJ, 97.

[32] Chen, J. Y. C. in press. Individual differences in human-robot interaction in a military multitasking environment. *J. Cognitive Engineering & Decision Making* (Special Issue on Improving HRI in Complex Operational Environments: Translating Theory into Practice).