THESIS

IMPLICATIONS FOR AUTOMATION ASSISTANCE IN UNMANNED AERIAL SYSTEM OPERATOR TRAINING

Submitted by

John G. Blitch

Department of Psychology

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2012

Master's Committee:

Advisor: Benjamin A. Clegg

Anne Cleary
Charles Anderson

ABSTRACT

IMPLICATIONS FOR AUTOMATION ASSISTANCE IN UNMANNED AERIAL SYSTEM OPERATOR TRAINING

The integration of automated modules into unmanned systems control has had a positive impact on operational effectiveness across a variety of challenging domains from battlefields and disaster areas to the National Airspace and distant planets. Despite the generally positive nature of such technological progress, however, concerns for complacency and other automation-induced detriments have been established in a growing body of empirical literature derived from both laboratory research and operational reviews. Given the military's demand for new Unmanned Aerial System (UAS) operators, there is a need to explore how such concerns might extend from the operational realm of experienced professionals into the novice training environment.

An experiment was conducted to investigate the influence of automation on training efficiency using a Predator UAS simulator developed by the Air Force Research Laboratory (AFRL) in a modified replication of previous research. Participants were trained in a series of basic maneuvers, with half receiving automated support only on a subset of maneuvers. A subsequent novel landing test showed poorer performance for the group that received assistance from automation during training. Implications of these findings are discussed.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

The ubiquitous arrival of computer based education in today's society dictates a prominent role for automation in the development of future training programs. The bloom (if not explosion) of internet connectivity at the turn of the century has given rise to a new field of Instructional Systems Design (ISD) which presents the potential for enormous cost savings in the form of self paced and personally tailored training in both interactive and automation enhanced simulation environments (Paquette, 2001). The escalation of automation enhanced training is also of enormous interest to the U.S. military and other government agencies that have to come to appreciate the value of unmanned system technology for a broad range of surveillance and reconnaissance activities. The Army for example has fielded thousands of tactical robots with tremendous variability in capability and Operator Control Unit (OCU) design that calls for extensive training transfer across platforms (Antal, 2009). The Air Force is striving to keep up with training demands for its most prominent Unmanned Aerial System (UAS), the Predator (Gramm & Papp, 2009). There is clearly a need to improve the effectiveness and efficiency with which unmanned systems operators are trained – leaving the door wide open for automation to be considered as an enabling mechanism.

The operational need for autonomous activity has been well established across a number of application domains (Canning, 2008; Carey & Markoff, 2010; Kurzweil, 2005; Sheridan, 2000; Singer, 2009b). As humanity endeavors to explore distant planets like

Mars, for example, control of robots and supervision of humans from earth will suffer extensive communications delays which prohibit the direct tele-operation modes that are common in the remote control (RC) hobbyist and entertainment communities (Sheridan, 1993; Sheridan & Verplank, 1978). Human fatigue and attention constraints also present strong motivation for remote surveillance and reconnaissance by autonomous robots in the context of lunar and Martian exploration(Crawford & Weisbin, 2005).

The case for automation in the application of mobile robot technology to the domestic emergency response domain is equally well established (Blitch & Maurer, 1996). Not only does the supervision of multiple robots combing an large search area imply that enormous value be placed on autonomous waypoint following, but the chaotic distribution of rubble makes radio controlled robots inherently vulnerable to communication interruption and failure. Such situations dictate the need for autonomous route planning of some sort to reestablish contact with human supervisors in pursuit of basic system reliability (Baker, Casey, Keyes, & Yanco, 2004; Micire, Drury, Keyes, & Yanco, 2009) .

Similar concerns regarding control loop latency and enemy jamming activities on the battlefield have inspired the military to develop unmanned systems with a high degree of autonomy (Antal, 2009; Hennigan, 2012; Singer, 2009a). The Army in particular has sought to invert the many-to-one ratio of unmanned systems currently operated by soldiers to a much more distributed paradigm where a single individual controls or supervises a number of platforms (Wickens, Dixon, & Ambinder, 2006). The U.S. Air Force apparently considers autonomous operation to be sufficiently important to the development of future battlefield robots that it will require full autonomous capability in its unmanned aerial systems to be implemented by 2047 (USAF, 2009). Such trends imply that the military will become increasingly reliant on recruitment and training of new UAS operators. One indicator of this change in the nature of UAS training is that the USAF has accepted non-rated pilots into its operator training program for the first time since initially fielding the Predator drone in the mid 1990s (Gramm & Papp, 2009).

In this context, the exploration of potential training shortcuts and enhancements gained through automated assistance presents a number of prominent advantages. Not only does automation reduce the cost and personnel assignment load associated with a large unmanned system instructional faculty, but it may also increase the intensity and density of training as students are allowed to practice critical skills on their own without dependence on instructor schedules and related resource allocation limitations associated with maintenance and operational requirements imposed by operational aircraft. Improvement in the nature of the training itself may also be realized through the use of automation to reduce counterproductive frustration levels associated with complex training tasks by allowing students to focus on enabling skills first before introducing them to more challenging integration tasks.

Operational automation concerns

Despite the prominent trend toward increased dependence on autonomy for unmanned systems control, however, there are a number of issues that provide ample cause for concern. These are evident on both an empirical and theoretical basis, particularly in high risk domains such as aviation safety and process control reliability which of course has dangerous consequences of its own in the context of nuclear power generation and other risk intensive activities.

Automation induced complacency.

Given recent events involving air traffic controllers who were fired for falling asleep on duty, perhaps the most salient concern at hand is the apparent complacency induced by automated systems which performed malfunction monitoring tasks too well. In two flight simulator experiments, for example, humans who were assisted with consistent malfunction alerts performed their overall vigilance tasks with significantly more error than those assisted with occasionally inconsistent reliability (Parasuraman, Molloy, & Singh, 1993). The apparent "cry wolf" nature of this effect has been replicated across a number of related tasks, indicating a robust yet deleterious impact that the aviation community in particular remains concerned about (Parasuraman, Sheridan, & Wickens, 2008; Wickens, 2009) in the form of increased safety violations and accidents due to ignored alarms.

Situational awareness loss – the "human out of the loop" problem.

Even in training situations where humans do not appear to commit an excessive number of errors during learning, any decrement in comprehensive understanding of current conditions and trends affecting the future (typically referred to as Situational Awareness or SA) cast in the shadow of automated assistance can have a potentially catastrophic impact on performance – particularly during emergency response activities that must be taken promptly and aggressively to prevent disasters (Endsley, 1997). In a sequence of prominent simulation studies validated with actual air traffic controllers on duty, Endsley and colleagues established a compelling case for the "human out of the loop" effect in which humans who have lost SA due to a reliance on automation may experience increased latencies in both problem detection & response (Endsley, 1997; Endsley & Kiris, 1995; Kaber, Onal, & Endsley, 1999; Kaber, Onal, & Endsley, 2000).

Because this loss of awareness does not always manifest in an immediate degradation of current performance, it can act as a kind of ticking time bomb which percolates just beneath a façade of safe operation, only to impede rapid response and resilience in the face of catastrophic consequences that might otherwise be avoided. In the most drastic cases, the response to urgent or dangerous situations is delayed or perhaps even obstructed until it is too late to avoid a crash or other disastrous situation (Endsley, 1997; Endsley & Kaber, 1999). For additional reviews regarding other forms of automation induced complacency see (Lee, 2006; Parasuraman & Riley, 1997; Sheridan, 1997; Wickens & Colcombe, 2007).

Automation influence on training.

Despite the abundance of empirical evidence regarding the potential for negative consequences of operational autonomy (Bailey & Scerbo, 2007; Bainbridge, 1983; Endsley & Kaber, 1999; Endsley & Kiris, 1995; Kaber, et al., 2000; Parasuraman & Riley, 1997; Parasuraman & Wickens, 2008; Smith, 2011), few have dealt with the autonomy decrement issue from a learning perspective other than through explorations of trust (Lee & Moray, 1992; Madhavan & Wiegmann, 2007; Sheridan & Parasuraman, 2000; Wickens, et al., 2006). Automation taxonomies emerging from a growing body of empirical evidence have established various Levels of Automation (LOA) following distinctions between failure detection (Sheridan & Parasuraman, 2000), functionality (Parasuraman & Wickens, 2008), and comprehensive perspectives which span the gamut of human machine interaction paradigms (Endsley & Kaber, 1999).

Since many of these studies have been conducted in professional environments involving air/ground traffic management and other high risk mission sets that preclude novice control due to safety concerns, the taxonomies on which they are based involve generally proceed on the important assumption that the humans involved have reached an appreciable degree of competence in the tasks at hand before automation is introduced. Although there is peripheral consideration given to autonomy influence on skill maintenance and refresher training, particularly in the closing years of the 20th century (Bainbridge, 1983; Moray, 1986; Parasuraman, Mouloua, & Molloy, 1996; Wiener, 1988), less consideration has been given to the impact of automation on the acquisition of skill itself.

It is with this relative paucity of literature dealing with the impact of automation on training that Clegg and colleagues launched their research into the influence of automation in the context of process control (Clegg, B.A., Heggestad, E.D., & Blalock, L.D., 2010). This effort explored the nature of autonomous assistance provided to novice operators who were trying to learn how to efficiently manage a moderately complex control process involved in the pasteurization of orange juice. Performance was measured on how much juice was successfully pasteurized in a given amount of time under full manual control and different levels of autonomous assistance made available under various conditions.

The results of this first experiment indicated an initial advantage presented by assistive autonomy. As training progressed and operators attained higher levels of proficiency, however, this advantage diminished to an insignificant level. A second objective of the study endeavored to explore any potential dependence that trainees might develop for the automation assistance. Removing automation during the final test phase impacted performance but only in the case where automation was automatically introduced during training.

By gradually decreasing the level of autonomy during training and randomly varying the specific control inputs managed by automation, a second experiment in this study sought to examine potential mitigation strategies which might be used to protect against autonomy induced learning decrements or perhaps even reverse them. The data, however, failed to support either approach. Not only was a performance decrement observed when the specific juice pasteurization subtasks were switched between manual and autonomous control on a random basis, but the automation removal effect persisted even when autonomy was gradually reduced over the course of instruction (Clegg & Heggestad, 2010).

By comparison, manual control performance during this study presented the strongest learning efficiency curve of all (on a performance increase over time basis), suggesting that while autonomy may present an initial advantage by projecting superior performance at the start of training (which may actually be quite important for novice trainee motivation levels), it appears to have a detrimental impact on training efficiency overall. While this might seem to be obvious in 20/20 hindsight and applied autonomy literature, the result at the time seemed quite counterintuitive from the perspective of novice trainees who appear to be overwhelmed by complex training tasks and might thus be expected to seek assistance from automated modules.

Maintaining a desirable level of difficulty in training has often been shown to keep such overconfidence tendencies in check while requiring a increased depth of processing (Bjork & Allen, 1970; Haider & Frensch, 2002; Healy et al., 2002; Healy et al., 2005) . Although increased difficulty does not always result in superior performance during learning (it often does quite the opposite), it has been shown time and again to enhance long term retention of acquired skill – which arguably constitutes the paramount goal of training in the first place (Bjork, 1994; Bjork & Bjork, 2006; Healy, Wohldmann, & Bourne, 2005).

Additional support for this notion can be found in the metacognition literature via the so called "testing effect" that shows time and again how simply reviewing information in a self study mode has been shown to be far inferior to testing one's knowledge in a more effortful yet fruitful (from a retention standpoint) retrieval paradigm (Carpenter & DeLosh, 2006; Kornell & Bjork, 2008). This begs the question as to whether autonomy may actually be helping too much in reducing the training task below a desirable level of difficulty that is necessary to form a comprehensive mental model of the task to be performed (Norman, D., 1990). Employees who have learned to rely on incomplete or inaccurate mental models developed during training may perform inadequately when it comes to overriding errant process control procedures – often with a disastrous consequence. Unfortunately, the impact of such a training deficit is destined to increase as the demand for innovative human problem solving is required at higher levels of complexity – a situation that has often been referred to "the irony of automation"(Bainbridge, 1983).

Despite these findings, it remains unclear what portion (if any) of automation influence on training observed in the process control regime would transfer into unmanned systems operator training. The taxonomy discussion put forth by Moray &

Inagaki suggests that these two task domains be considered differently (Moray & Inagaki, 1999). Observations of participants learning air combat management tasks during the advent of computer enhanced cockpits also suggest that these two task domains be considered differently. Ballas and colleagues observed that participants in a flight simulator performed combat management (a rough analog to what is considered "process control") tasks differently (i.e. with occasionally better performance via text selection) during flight operations than tactical maneuvers performed via direct manipulation of an aircraft's control surfaces via joystick (Ballas, Heitmeyer, & Pérez-Quiñones, 1992).

The goal of the experiment that follows was to replicate the automation effects observed by Clegg and colleagues (Clegg, Benjamin A., Heggestad, Eric D., & Blalock, Lisa Durrance, 2010) in the context of unmanned systems control. Although this previous work established automation induced decrements in the training regime, it remains unclear whether those effects were specific to the process control domain or not. If the distinction between manual dexterity and cognitive skill that Bainbridge notes in describing process control oscillation can be mapped to the tactical cockpit as Ballas and colleagues suggest, then the manner and intensity with which automation influences these two different skill sets may vary accordingly (Bainbridge, 1983; Ballas, et al., 1992).

In any case, the intent here is not only to examine the consistency of the automation withdrawal effect in a context with relevance to unmanned systems operating in a military setting, but to do so in a domain that is exciting and challenging enough to avoid the potential apathy and boredom induced by a typical keyboard/mouse input paradigm (Gee, 2003). By adding a complex motor control challenge applied to an unfamiliar (and notoriously difficult) aircraft flight task, participants were expected to welcome assistance from the autopilot much more warmly than a vigilance dominated process control task.

Since this flight simulator emulated some of the typical video games in today's entertainment market, this experimental design endeavored to minimize the role that motivation plays in training as explored in previous research (Barab et al., 2009). Because the gaming paradigm presents a more entertaining task set than the pasteurization process, a relatively constant level of engagement was anticipated across the autonomy and manual groups and thereby emphasizing a task difficulty manipulation while holding motivation relatively constant in deference to previous research (Broadhurst, 1959).

Another factor leading to the current design was the intent to invoke a standardized comprehensive task training procedure in lieu of the trial and error style of learning used by Clegg et al (2010). This more structured approach was expected to avoid inconsistent oscillations from extraneous factors occasionally observed in the trial and error process which might otherwise have allowed automation to fundamentally shift the level of task understanding being developed for the entire system. The standardized process, by comparison, was expected to limit the impact of automation to only influence the experiential portion of learning rather than all possible aspects of the trial and error process.

The design used here also included more a common experience between the two control groups since automation was only introduced in a middle training block sandwiched between two manual control blocks. By starting and ending with a common training mode, any negative consequence of automation withdrawal (such as motivation or attentional anomalies) that did not have direct learning implications for the task at hand were reduced. This consistency factor is of particular importance in considering that the final test, landing the aircraft, represented a culminating yet novel challenge which required trainees to integrate the skills they learned during the three component sub tasks in a new way.

This approach allowed us to examine how automation influenced the trainees' understanding of the underlying principles of unmanned systems control rather than just the physical relationships associated with routine flight characteristics. In combat environments where anomalous activity and situational dynamics demand that emergency procedures be invoked well beyond what routine relationships might otherwise handle, a comprehensive understanding of the entire systems control realm is crucial to successful recovery from otherwise disastrous circumstances whether they are induced and/or exacerbated by automation, enemy activity, or any number of other factor combinations.

By utilizing the U.S. Air Force simulation based training paradigm for novice Predator operators, these issues were addressed enroute to the primary objective of the research – to examine the hypothesis that automation induced deficits previously observed in the process control regime extend into the realm of unmanned systems control, and thereby warrant research and development of mitigating strategies to counter their negative influence.

CHAPTER 2: EXPERIMENT

The following study was conducted in order to assess the potential for autonomy assisted training to increase or decrease the learning efficiency of novice trainees struggling to learn a highly complex task in a militarily relevant domain. The primary goal was to test the autonomy removal hypothesis that predicts a drop in overall performance when trainees who have used assistive autonomy are subsequently required to perform a complex task without it. By requiring one group to train manually while providing another group access to assistive autonomy (in the form of a selective auto-pilot function in Predator UAS simulator) during the middle of three training blocks, the current experiment aspired to compare performance levels in a manner similar to the orange juice process pasteurizer process used by Clegg et al. (20102), but in a more complicated task where autonomy might provide a welcome relief from overwhelming complexity.

Method

Participants

20 Colorado State University (CSU) undergraduates participated in this study for optional, partial, or introductory psychology course credit. Participants were randomly assigned to one of two groups: one requiring manual control during the entire training and test process, and another that was provided with autonomy assistance during one phase of their training.

Materials

This experiment was conducted using the Predator STE (Synthetic Training Environment) simulation software provided by the Air Force Research Laboratory (AFRL) installed on a Dell Pentium 4 desktop computer with a Hands On Throttle And Stick (HOTAS) joystick assembly and secondary monitor (USAF/AFRL, 2002). Input stimuli of interest were provided by the Predator STE simulator as described by Martin and colleagues in their report on development of synthetic task for human factors development (Martin, Lyon, Schreiber, & Martin, 1998). A pictorial representation of this basic setup is included in Figure 1.

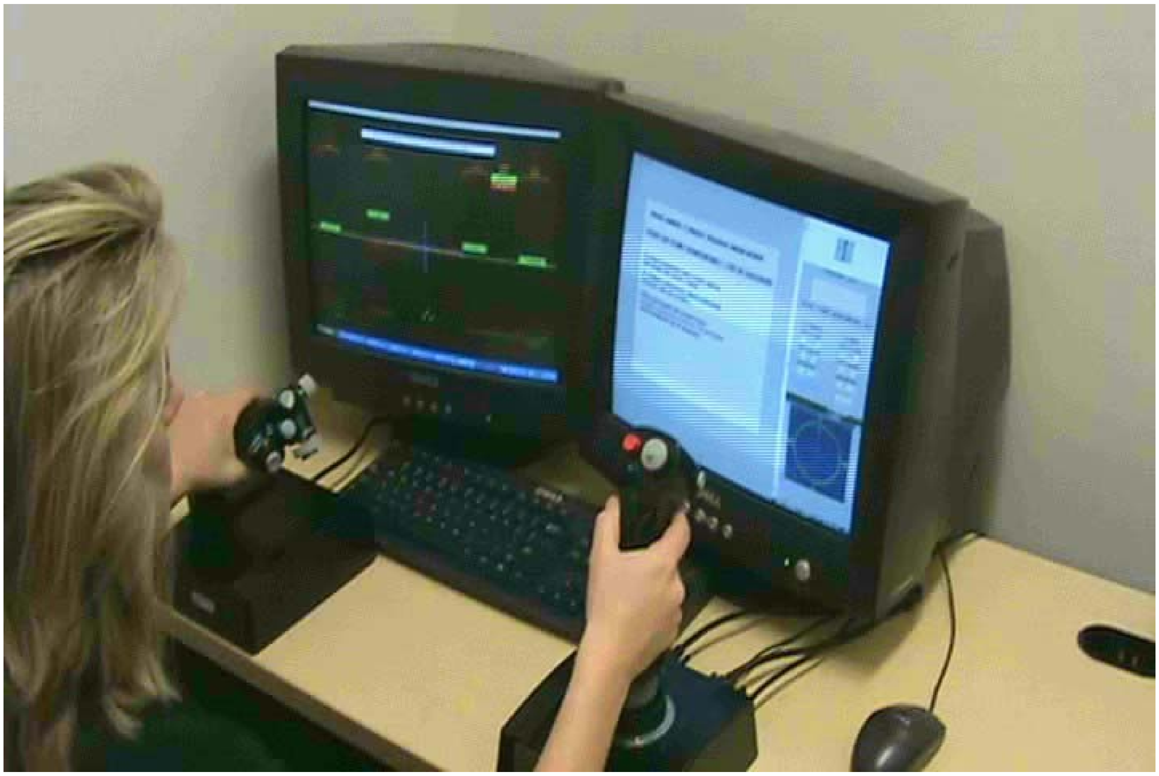

Figure 1. Predator Synthetic Training Environment (STE) basic setup

Procedures

This experiment used a between subjects design with a manual (M) group and autonomy assisted (AA) group. Participation in this experiment was conducted in four sessions each lasting approximately one hour in duration. Prior to initiating hands on training blocks, participants were asked to review a 25 minute tutorial on the Basic Maneuver (BM) characteristics of the Predator UAS. This tutorial was comprised of a typical slide show type presentation with animated video clips to provide moving representations of the more complex aspects of aerodynamic flight and aircraft control surfaces.

The first hands on training block (TB1) commenced with the participant reading over a one page scenario description of the STE's basic maneuver scenario one. This TB1 task required them to reduce airspeed from 67 knots down to 62 knots while holding heading and altitude constant. The participant was given five minutes to read the page and ask for clarifications.

The participant then performed 20 training trials on TB1, each lasting one minute in length. Performance for each trial was automatically recorded and displayed for the participant at the end of each trial in the form of their Root Mean Square Error (RMSE) compared to optimal flight control inputs as indicated by the feedback panel in Figure 2. Participants were typically required to take a brief (approximately one minute) break after every 5 trials in order to mitigate fatigue effects. This break was also used as an opportunity for experimenters to check data recordings for consistency and accuracy.

In order to avoid instructional bias, experimenters offered no coaching or advice of any sort. Questions from participants regarding functionality of any sort were answered with an offer to conduct a brief review the previously shown tutorial as a refresher. Subsequent questions following that procedure were simply answered with "please just do the best you can." Sessions were scheduled on four consecutive days starting on Monday and ending on Thursday whenever possible to maintain a standardized training profile. This process also ensured a consistent opportunity for sleep enhanced memory consolidation between training blocks.

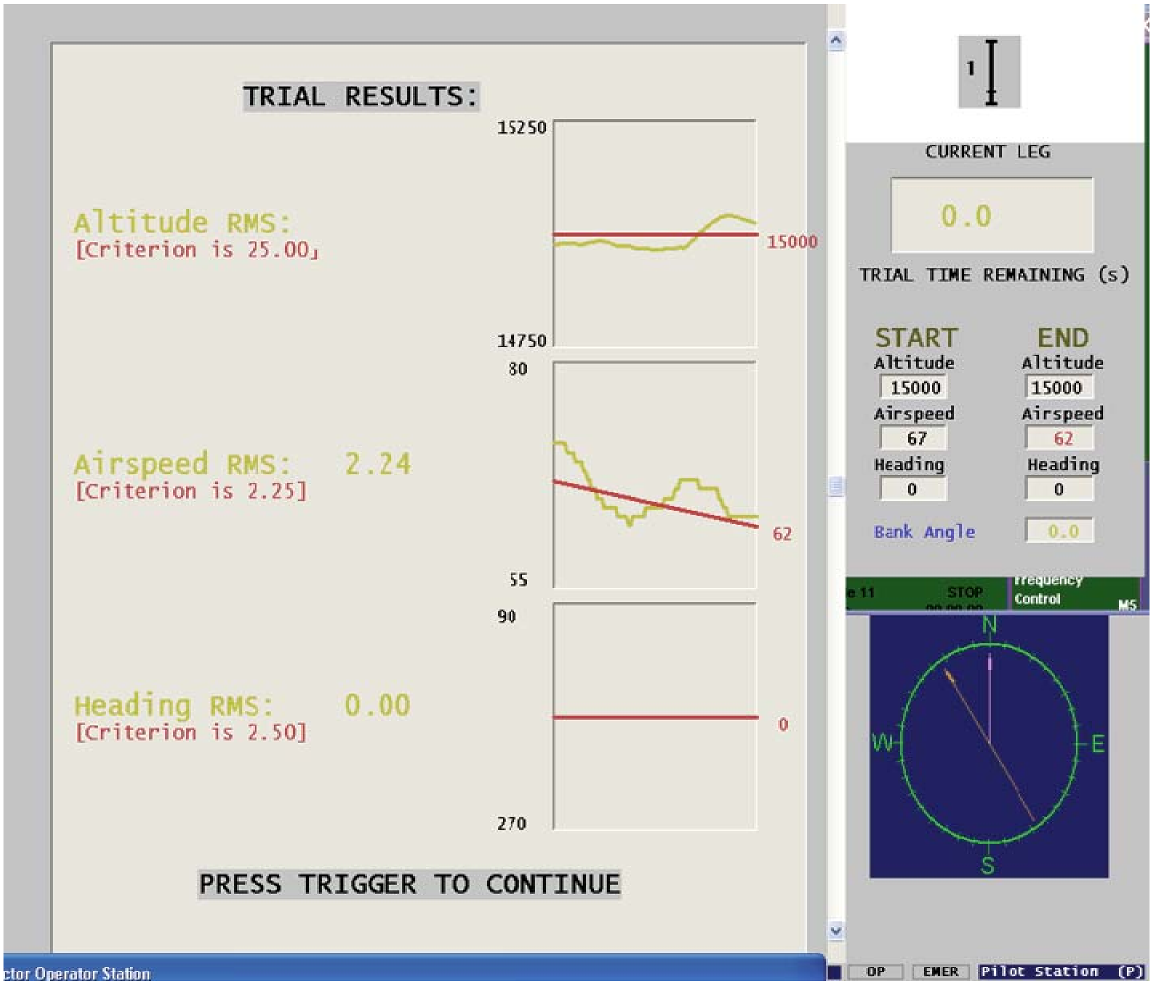

Figure 2. Example of performance feedback during Basic Maneuver training

Training block 2 (TB2) began by having participants read a single page description of the Predator STE's basic maneuver scenario two. This TB2 task required participants to execute a 180 degree turn while holding altitude and airspeed constant. Participants who had been randomly assigned to the manual group were provided with scenario description TB2M (for Manual control) which contained standard instructions for maintaining altitude and airspeed during the turn. Those who had been randomly assigned to the autonomy group were provided with the scenario description of TB2A (for Automation assistance) which explains that both the throttle and pitch inputs during this turning scenario would be handled by the aircraft's autopilot. Thus the sole manual input for the automation assisted group was along the roll axis, controlled via the joystick. These participants were therefore instructed to ignore the performance feedback for altitude and airspeed and focus on heading instead. Participants were once again allowed a brief stretch break after every five of their twenty trials.

At the beginning of Training block 3 (TB3) participants were provided with a one page description for the STE's basic maneuver scenario five which requires a straight line descent (reduction of airspeed and altitude while holding heading constant) under full manual control.[1] The final test was focused exclusively on a descent to land paradigm and involved no climbing activity of any sort.

The final session of the experiment constituted the test block and began with each participant completing the Predator STE's Landing Task (LT) tutorial in the same fashion as the basic maneuver tutorial. Once the LT tutorial was been completed, each participant was provided with a single page scenario description for the Land Task Test (LTT) which required them to perform two 90 degree descending turns while on approach to the runway, and then land the aircraft. They performed this task five times, and reviewed their performance data after each trial with the feedback indicated in Figure 3.

LTT trials typically lasted from three to four minutes depending on the approach and landing path taken. After all five landing trials were completed the participant was asked to fill out a short (five minute) spatial experience survey which allowed experimenters to distinguish between true novice trainees and those with actual flight

[1] It should be noted that the climbing and aircraft ascent tasks included in basic maneuver scenarios three and four were omitted since they fell outside the scope of the transfer test - landing the aircraft.

experience or other exceptional experience that might result in anomalous training effects. This survey verified the novice status of all participants in the study.

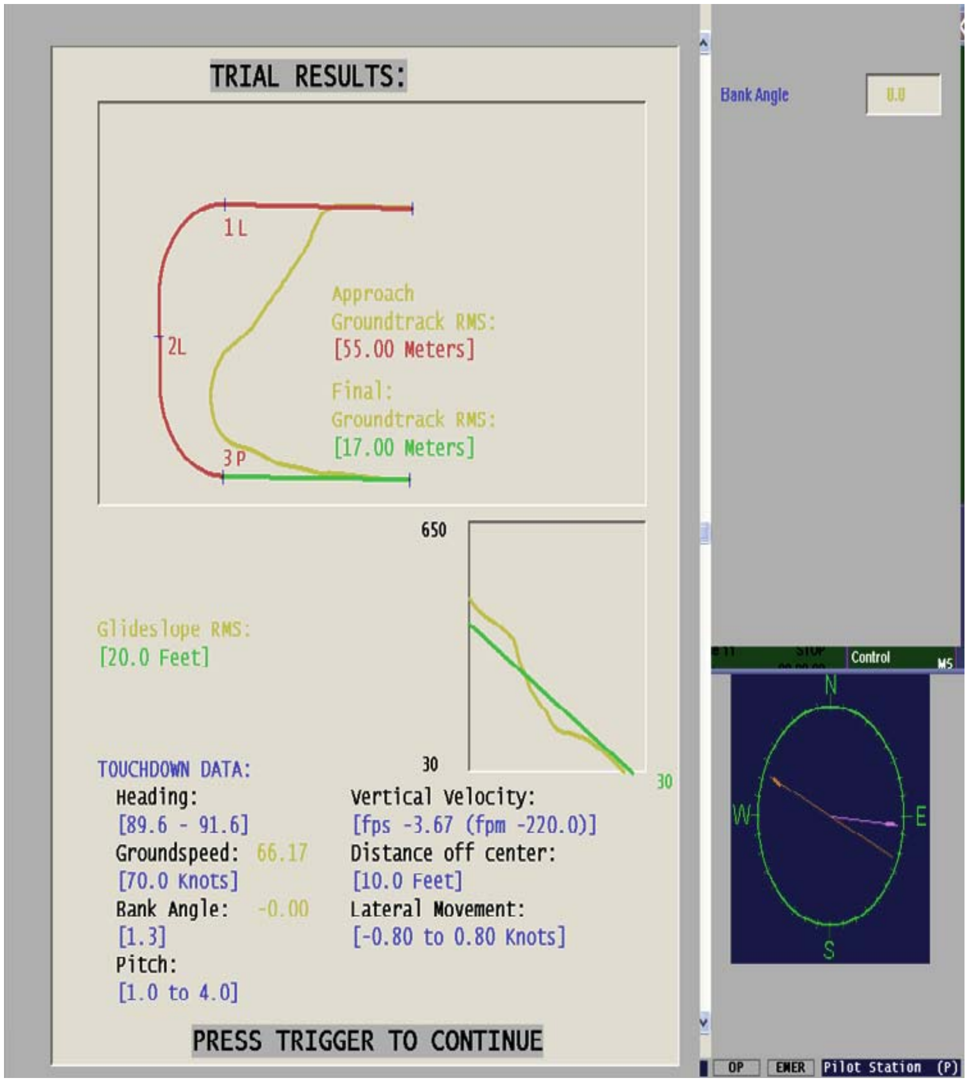

Figure 3. Example of performance feedback during the Landing Task Test

Measures

The Predator STE simulator records individual performance for every trial. In this experiment, the data collected during basic maneuver sessions was considered to be training, while landing task records were considered to be test data.

Performance metrics during training.

The Predator STE simulator recorded performance during training in the form of error metrics which represented the aircraft's response to the primary control inputs that are required for an operator to successfully maneuver and ultimately land the Predator aircraft. These automatically calculated by the simulator's path projection modules in a Root Mean Square Error (RMSE) fashion captured in the 3 following metrics:

Altitude error was measured as a y axis vertical displacement from the optimal path that a trainee should have taken on each trial as indicated by instructions given at the introduction for each task. In the first basic maneuver task (TB1), for example, the trainee was required to reduce airspeed while holding altitude constant at 15,000 feet. Altitude error during this task was measured as the absolute value above or below 15,000 that the trainee allowed the aircraft to climb or descend over the course of each trial.

Heading error was measured as a bearing difference in degrees measured from the optimal direction that a trainee should have maintained in relation to the path specified in each task scenario. In the second basic maneuver task (TB2), for example, the trainee was required to execute a smooth 180 degree turn at an optimal rate of three degrees per second. Heading error during this task was measured in degrees from the dynamically adjusted direction required to maintain a smooth 180 degree arc over course of each trial.

Airspeed error was measured as the difference from the optimal Indicated Air Speed (IAS) required to accomplish the task at hand. During TB2, for example, the trainee is required to perform a 180 degree turn while maintaining airspeed at constant of 62 knots. Indicated Air Speed (IAS) error during this task was measured as the difference above or below that constant airspeed of 62 knots that the trainee allowed the aircraft to achieve during each trial.

Criterion Pass/Fail: In addition to the error metrics listed above, the number of times a participant reached a criterion performance level (automatically established by the simulator and indicated to the trainee during post trial feedback) was also recorded as an indicator of how much learning was accomplished by each participant.

Although these error measures are related to the ground track and glide slope metrics later considered under the general rubric of the Landing Task Test, they do not constitute a direct mapping to landing proficiency or even control inputs themselves. In many cases, the specific error they endeavor to capture reflects a summation of multiple control inputs. Failure to maintain a specified altitude, for example, can result from having the aircraft pitched up or down too much through overly aggressive manipulation of the joystick and/or simultaneously applying too much or too little throttle input to the engine - perhaps while struggling to maintain the appropriate bank angle required to execute a smooth turn.

Maintaining a proper balance between the three primary control inputs: pitch, roll, and throttle (yaw/rudder control was fully automated by the simulator's default settings), represents the essence of the control challenge for those learning to land the aircraft. If the trainee was able to maintain a balance of control inputs which kept the aircraft within an acceptable performance criterion envelope for each metric during an entire trial, then simulator would automatically record a "pass" rating for that particular task/scenario (TB1 for example) and the participant would be started on the next successive task/scenario (TB2).

In order to ensure that trial by trial training time for each participant was held constant, trainees were restarted on the same scenario for all 20 trials regardless of whether or not they passed criterion on that particular run. Thus all participants completed all 20 trials of each training block regardless of the level of competence they ultimately attained.

Performance metrics at test.

Final performance in this experiment was automatically calculated by the simulator in the RMSE fashion indicated above and captured in 3 separate metrics as described below:

Approach Ground Track error was measured as an x/y horizontal displacement from the optimal approach path that a pilot candidate should have pursued in relation to various terrain features in the vicinity of the airport while turning the aircraft through three waypoints enroute to their final landing activity.

Final Ground Track error was measured as an x/y horizontal displacement from the optimal approach path that a trainee should have taken in relation to the runway once they had made their final turn toward the airport and were on the final leg of their landing.

Glide Slope error was measured as a +/- vertical displacement along the z axis of an optimal descent slope that the trainee should have pursued through each LTT trial from start to finish (touchdown).

Criterion Pass/Fail: In addition to the error metrics listed above, the number of times a participant reached criterion was also recorded as an indicator of how much learning was accomplished by each participant. This is the only performance metric that was common to both the basic maneuver training phase and the landing task test phase, although it was calculated differently for each.

CHAPTER 3: RESULTS

Performance results at test

In order to answer our research question regarding the influence of automation on learning, performance on the Landing Task Test (LTT) was initially investigated as the primary indicator of training efficiency since hands on duration, or "stick time" was held constant for all participants. This analysis was focused on mean performance measured across the three primary landing task metrics indicated above.

In order to explore test performance beyond the potential influence of first trial nervousness, a value of average error across all five trials was compiled for each test metric and subjected to an independent samples t-test. This resulted in a statistically significant difference in glide slope error indicating that the Automation Assisted (AA) group ($M=186$, $SD=60$) performed worse that the Manual (M) group ($M=124$, $SD=65$) in handling altitude and airspeed while landing the aircraft $t(18)= 2.24$, $p<.05$. These data presented a relatively large effect size (Cohen's $d =0.91$) as depicted in Figure 4.

No group difference was found in the Approach Ground Track metric (AA group $M=123$, $SD=59$, M group $M=103$, $SD=34$), $t(18)= 0.92$, $p>.05$, nor the Final Ground Track metric (AA group $M=100$, $SD=65$, M group $M=85$, $SD=67$),$t(18)= 0.51$, $p>.05$). But importantly, neither showed a trend towards an advantage of automation in training – suggesting that no benefits were derived from allowing the AA group to focus on only one portion of the control requirement when automation was present.

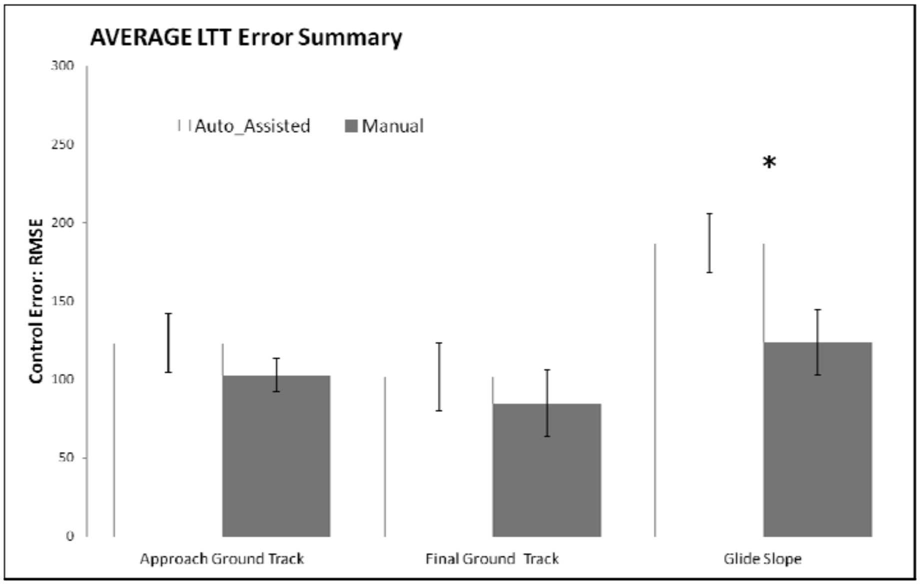

Figure 4. Average performance across all five Landing Task Test trials for groups that were previously assisted with automation or trained solely with manual control

After this initial difference in average Glide slope error was found, an inspection of means from the first Landing Task Test (LTT) trial was conducted, showing that while a moderately larger amount of error appeared in the data graphs for the Automation Assisted (AA) group (M=283, SD=142) compared with the Manual (M) group (M=195, SD 101) in Figure 5, this trend failed to reach statistical significance t(18)= 1.59, p>.05., Cohen's d =0.69.

No significant difference in first LTT trial error was found in the Approach Ground Track metric (AA group M=138, SD=125, M group M=160, SD=128), t(18)= 0.39, p>.05, nor the Final Ground Track metric (AA group M=146, SD=214, M group M=124, SD=186), t(18)= 0.24, p>.05.

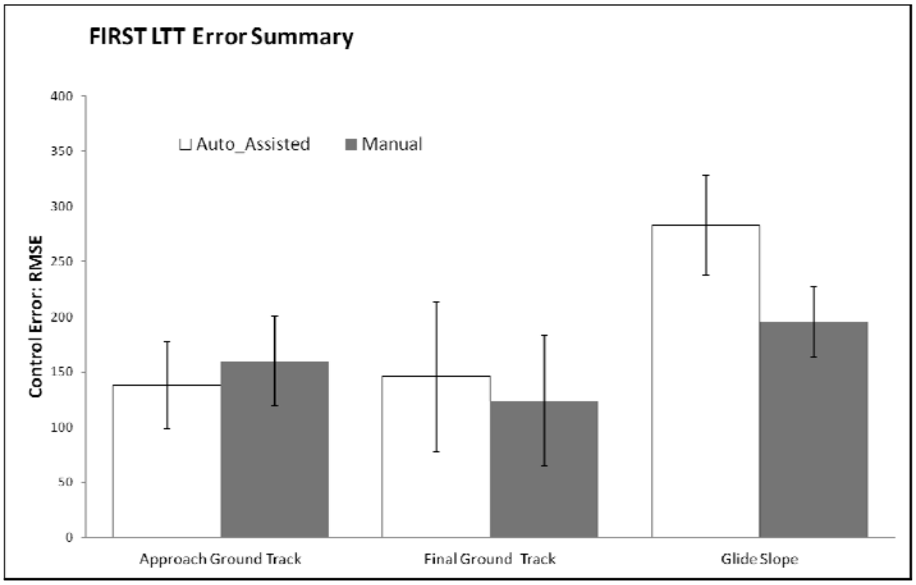

Figure 5. Average performance results from Landing Task Test across all 5 trials

An identical analysis of the data collected on the last LTT trial failed to show a statistically significant difference between groups ($t(18)= 0.63$, $p>.05$, Cohen's $d=0.28$) suggesting that the AA group ($M=132$, $SD=64$) was able to reduce their relative deficit in Glide Slope control during landing compared to the manual group ($M=111$, $SD=80$) as indicated in Figure 6. No difference in last LTT trial error was found in the Approach Ground Track metric (AA group $M=89$, $SD=41$, M group $M=78$, $SD=43$), $t(18)= 0.58$, $p>.05$, nor the Final Ground Track metric (AA group $M=71$, $SD=75$, M group $M=74$, $SD=64$), $t(18)= 0.07$, $p>.05$).

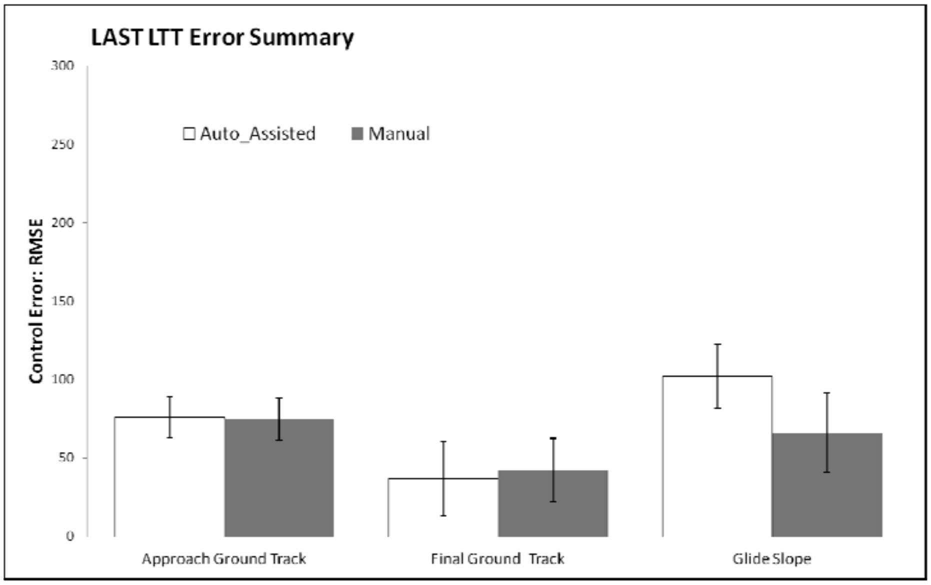

Figure 6. Performance results from final Landing Task Test trial

In order examine the notion that the AA group might have been able to recover from their initial performance deficit at the start of the Landing Task Test, a repeated measures Analysis of Variance (ANOVA) was conducted across all five landing attempts on trial by trial basis.

Analysis of the Glide slope data was found to violate Maunchly's test for sphericity (X2(9) = 25.727, $p<.05$), so the Greenhouse-Geisser correction (€ = .720) was used. After this correction was made, a main effect for trial ($F(2.882, 4)= 4.428$, $p<.05$, $\eta_p= 0.174$), showed that performance was changing across LTT trials. A significant test of linear contrast ($F(1)=10.888$, $p<.05$, $\eta_p= 0.341$), indicated that performance showed consistent changes across trials. A main effect for group was also found ($F(2)=5.054$, $p<.05$, $\eta_p= 0.325$) but there was no evidence of an interaction of group with trial ($F(5.764, 8)= 0.438$, $p>.05$) as portrayed in figure 7.

Taken together, these analyses indicate that although both sets of participants were able to steadily reduce their error over time, the automation assisted group consistently performed worse in controlling glide slope during landing than the manual group. While a comparison of effect size suggests that the automation assisted group was able to reduce the relative impact of their glide slope control deficit over time, the lack of an interaction between group and trial discourages any claim of slope convergence that might otherwise suggest that the AA group was essentially able to catch up to the M group due to continued learning during the test phase. Further examination of this convergence potential is left as motivation for future research.

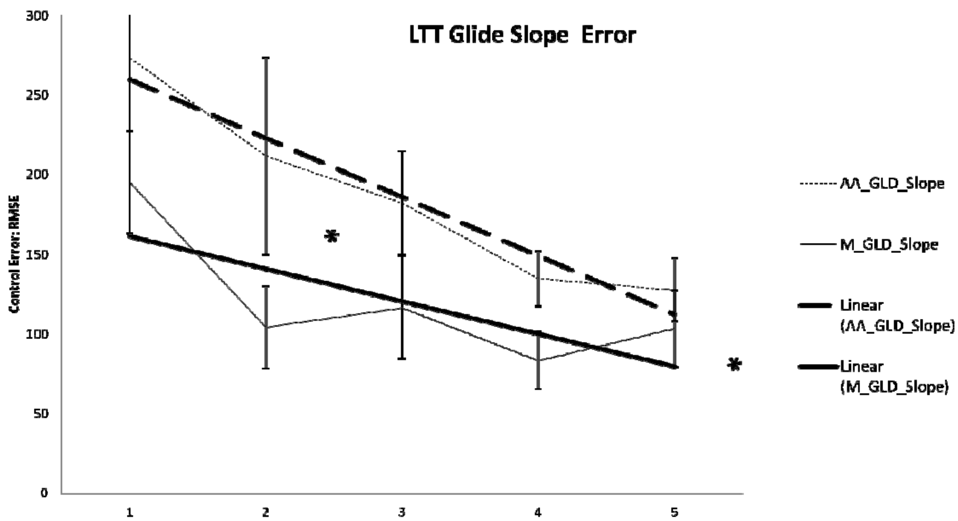

Figure 7. Trial by trial Glide slope error for Landing Task Test

Approach Ground Track data was also found to violate Maunchly's test for sphericity ($X^2(9) = 35.13$, $p<.05$), so the Greenhouse-Geisser correction ($\epsilon = .504$) was again used to correct degrees of freedom. After this correction was made, a main effect for trial ($F(2.014, 4)= 4.153$, $p<.05$, $\eta_p= 0.165$) showed that performance was changing across LTT trials. A significant test of linear contrast (F(1)=7.218, p<.05, $\eta_p= 0.189$), again indicated that performance showed consistent changes over the course of the test block. No main effect was found for group in the Approach Ground Track metric, however (F(2)= 2.454, p>.05, η_p= 0.189), nor was there any interaction observed between trial and group (F(4.028, 8)= 0.814, p>.05, η_p= 0.072). This analysis suggests that both groups managed to decrease their Approach Ground Track error steadily with each additional trial as indicated in Figure 8.

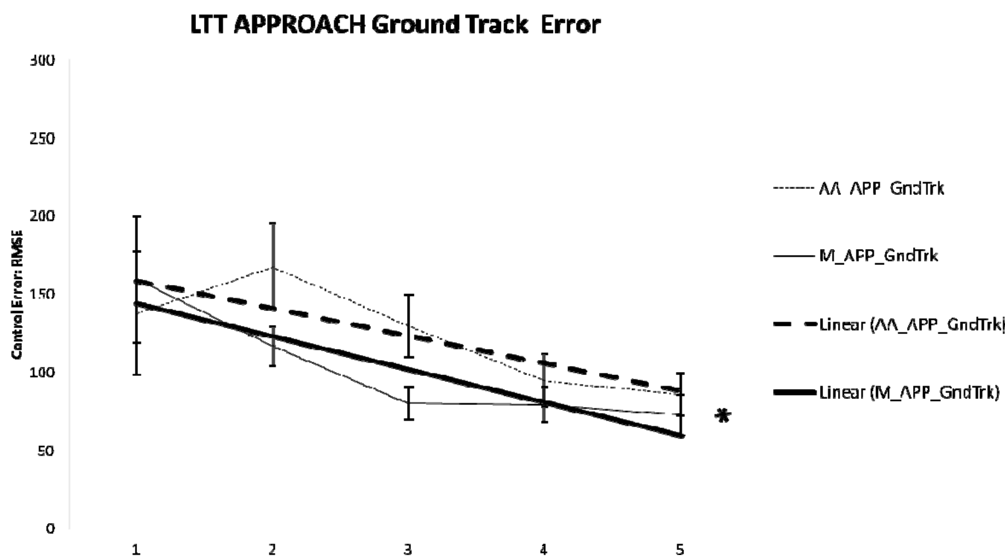

Figure 8. Trial by trial Approach Ground Track error for Landing Task Test

Final Ground Track error was analyzed in an identical manner to the other performance metrics above. Mauchly's test for sphericity was again violated ($X^2(9)$ = 53.37, p<.05), so a Greenhouse-Geisser correction (ϵ = .459) was again used to correct degrees of freedom. After this correction was made, no main effect was observed for trial (F(1.837, 4)= 1.513, p>.05, η_p= 0.067) or group (F(2)= 0.884, p>.05, η_p= 0.078), nor was there any evidence of an interaction between the two (F(3.674, 8)= 0.870, p>.05, η_p= 0.077) as illustrated in figure 9. This data presents no evidence that learning occurred on this control variable across these particular trials, perhaps due to a floor/ceiling effect or other influence that is too gradual to detect.

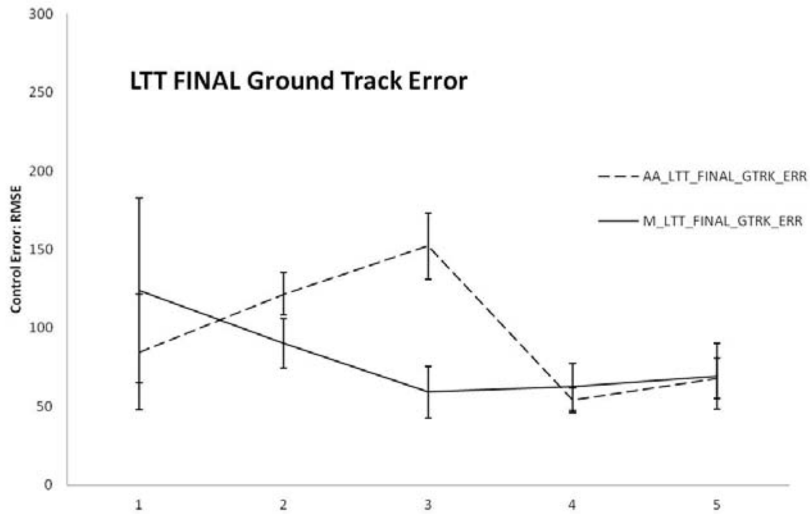

Figure 9. Trial by trial Final Ground Track error for Landing Task Test

It should be noted here that no participant ever reached criterion on the Landing Task Test. Not only does this show how difficult the Predator aircraft is to land effectively, but it also indicates how early these trainees are being assessed in what might otherwise be a professionally mandated training cycle. Exploration of extended training efforts that might allow participants to reach landing task criterion remains an issue to address with future research.

Performance results during training

While the skills transfer test from basic maneuvering to landing offer the best examination of the hypotheses at hand, questions remain about the impact that automation had on the nature of skill acquisition itself. In order to comprehensively assess the impact of automation on training, it makes sense to expand analysis beyond test data and examine whatever learning indicators might exist within the training data itself. With this objective in mind, a secondary analysis was launched which sought to take advantage of the automatic error recording features available in the Predator STE simulator. This analysis followed the same general rubric as the test analysis, but with a more direct focus on control inputs indicated by the training performance metrics: altitude, heading and airspeed error. The number of times each participant was able to meet criterion was also considered on a pass/fail basis over the total of 20 trials per training block.

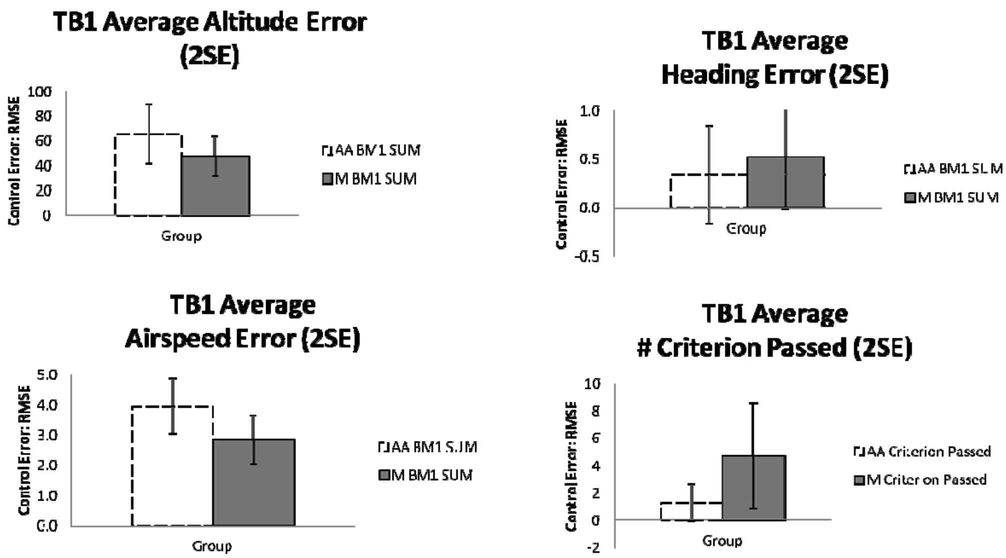

Figure 10. Average error summary for Training Block 1

Although the functional distinction between the Automation Assisted (AA) group and the odd numbered Manual (M) group was not made until the autopilot function was actually introduced during training block 2 (TB2), data from training block 1 (TB1) was still evaluated to determine whether or not the assumption of randomly assigned participants (via even and odd numbered enrollment numbers respectively) was valid. As expected, a series of independent sample t-tests conducted on average error data from TB1 revealed no significant difference between groups in altitude error AA ($M=66$, $SD=38$), Manual ($M=48$, $SD=25$), $t(18)=1.26$, $p>.05$, airspeed error AA($M=4.0$, $SD=1.5$), M(M=2.8, SD=1.2), t(18)=1.86, p>.05, or heading error AA(M=0.34, SD=0.80), M(M=0.52, SD=0.80), t(18)=0.52, p>.05 as indicated in Figure 10.

The data for criterion achieved during TB1 failed Levene's test for equality of variances. After a correction to the degrees of freedom was made, the two tailed independent samples t-test (t(11.2, 18)=2.15, p>.05) revealed no difference in pass/fail criteria between the Automation Assisted group (M=1.3, SD=2.2) and the Manual group (M=5.7, SD 6.1) also indicated in Figure 10.

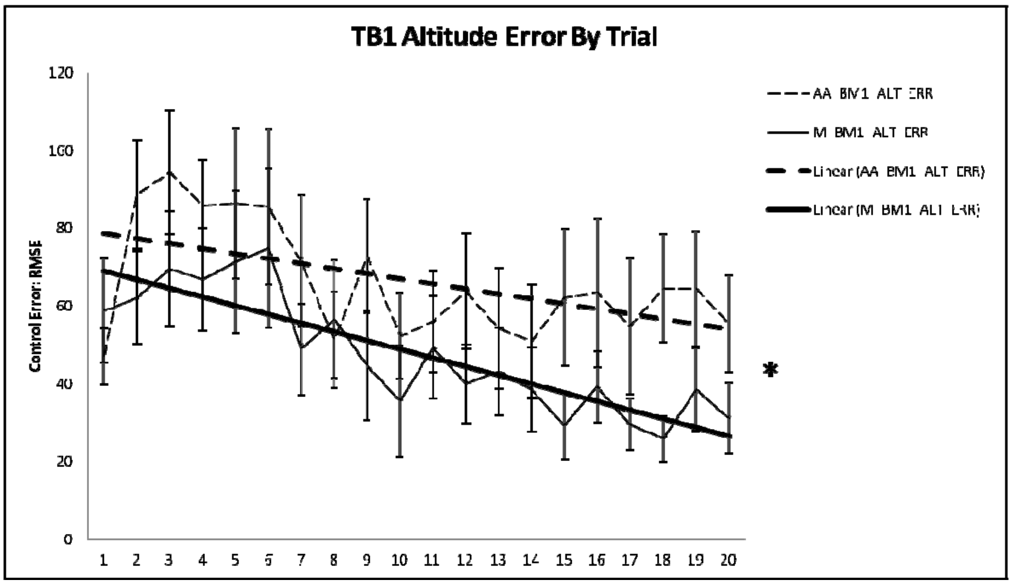

Figure 11. Trial by trial altitude error during Training Block 1

Following the precedent set on above for the landing task, a trial by trial analysis for TB1 was conducted via repeated measures ANOVA for each of the training error metrics. Mauchly's test of sphericity was consistently violated during this analysis, so a Greenhouse-Geisser correction was applied throughout. A main effect and linear fit for TB1 trials was found for altitude (ϵ = .293), (F(5.570, 19)= 3.836, p<.05, η_p= 0.176), (F(1)=10.659, p<.05, η_p= 0.372) and airspeed, (ϵ = .201), (F(3.816, 19)= 6.816, p<.05, η_p= 0.275), (F(1)=12.432, p<.05, η_p= 0.409) indicating that accumulated error was changing significantly and consistently during training on these two metrics, but not for heading (ϵ = .122), (F(2.318, 19)= 0.857, p>.05) as portrayed in Figures 11-13. These data suggest that participants were learning how to reduce their control error in altitude and airspeed but not heading, which is consistent with the apparent objective of this particular training block.

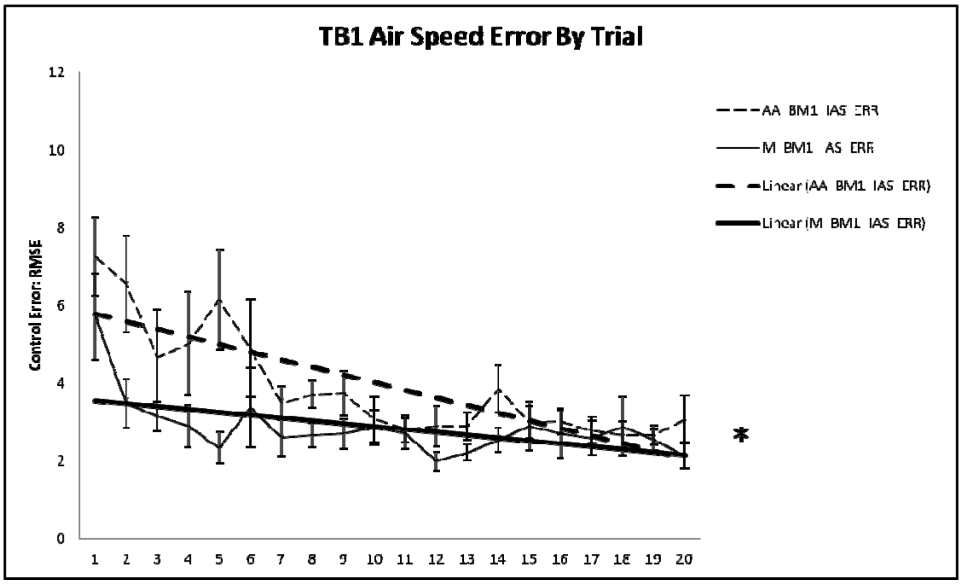

Figure 12. Trial by trial airspeed error during Training Block 1

Given the TB1 scenario requirement to hold heading and altitude constant while reducing airspeed, these data make sense because the amount of heading error incurred via side pressure on the stick while participants endeavored to compensate for slower speed with increased pitch was negligible. There was, in essence, no significant source of heading error to be reduced in this training scenario.

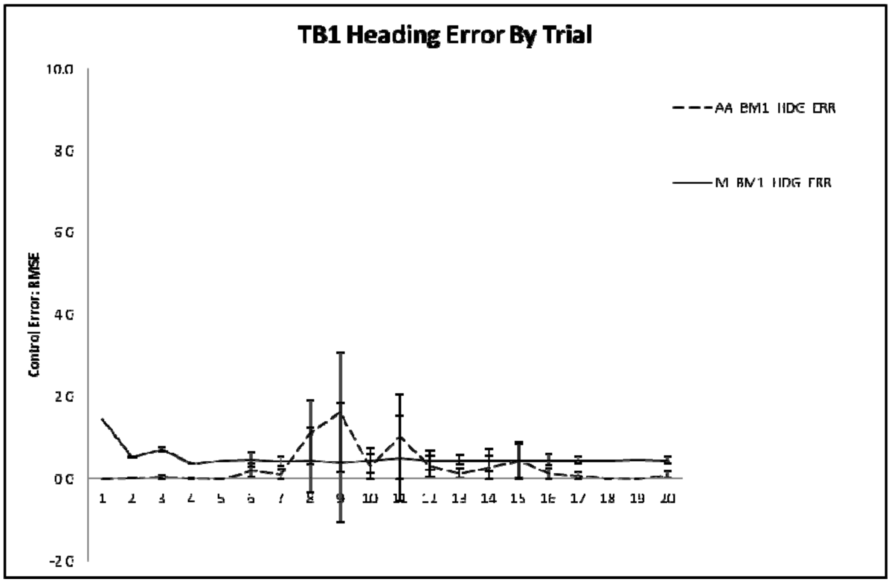

Figure 13. Trial by trial heading error during Training Block 1

Importantly, there was no main effect for group or interaction between trial and group evident during TB1 for altitude $F(1)= 1.714$, $p>.05$), (Є = .293), ($F(5.570, 19)= 0.755$, $p>.05$, airspeed $F(1)= 3.328$, $p>.05$), (Є = .201), ($F(3.816, 19)=1.796$, $p>.05$) or heading $F(1)= 0.449$, $p>.05$), (Є = .122), ($F(2.318, 19)= 1.203$, $p>.05$), which suggests that all participants started out with equivalent performance before automation was introduced.

When considered collectively, these data illustrate that both groups of participants were able to steadily decrease their altitude and airspeed error in a linear fashion over the 20 trials in the first training block, which suggests that an appreciable amount of learning took place without any significant difference between groups. This notion was further supported by the average number of times (M=3.50, SD=4.99) that all participants were able to reach criterion in TB1 no matter which group they were assigned to.

An analysis of training performance during the second training block (TB2) could only be compared via the heading error metric since airspeed and altitude were controlled by the autopilot for half of the participants. No difference was found between groups in a t-test comparison of average heading error AA(M=29.5, SD=14.6), M(M=24.8, SD=12.7), t(18)=0.769, p>.05, as illustrated in Figure 14.

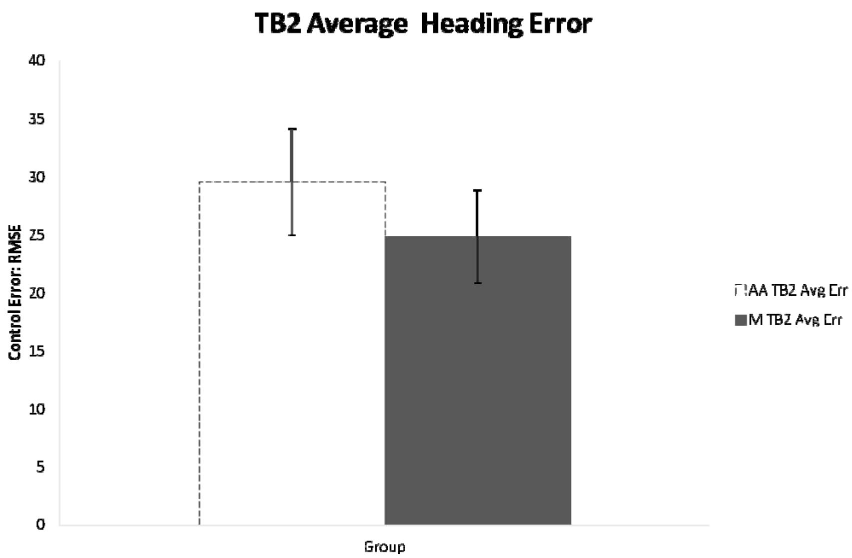

Figure 14. Average heading error during Training Block 2

Although participants under full manual control were able to achieve criterion on an average of 1.1 of their 20 trials in TB2, it would be inappropriate compare that number against a zero value for the AA group since the imperfect nature of the STE autopilot prevented participants receiving automation assistance from reaching criterion on any trial.

A sequential analysis of TB2 heading data revealed a main effect for trial ($F(5.409, 19)=11.979$, $p<.05$, $\eta_p=0.400$, with a significant linear contrast $F(1)=39.178$, $p<.05$, $\eta_p=0.685$ after a Greenhouse-Geisser correction ($\epsilon = .285$) for sphericity violation was applied. There was no main effect found for group $F(1)=0.591$, $p>.05$, and no interaction of group with trial ($F(5.409, 19)=1.083$, $p>.05$) as indicated in Figure 16.

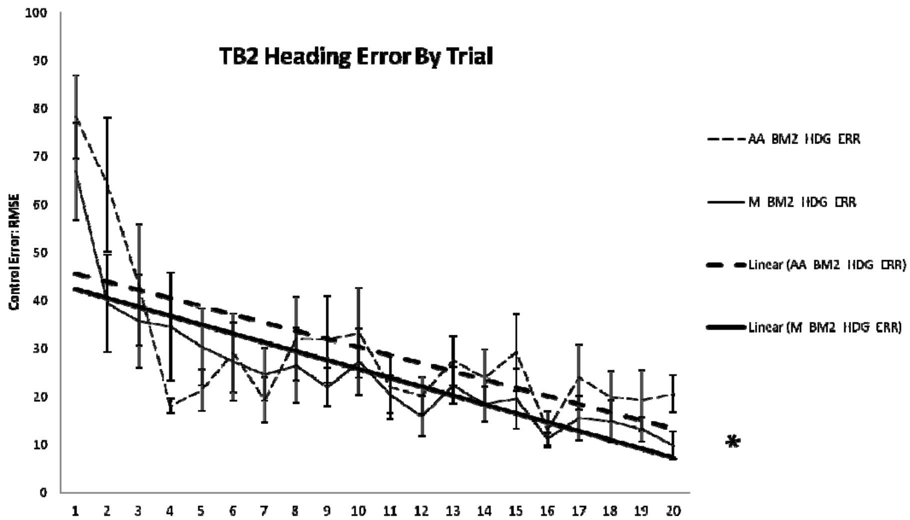

Figure 15. Trial by trial heading error during Training Block 2

These data were surprising in several respects. While the linear reduction of error during TB2 was generally expected commensurate with the power law of practice (Newell & Rosenbloom, 1981), the lack of a significant error difference between groups was not. Given that two of the three control inputs required during the TB2 turning task were handled by the autopilot for the participants in the AA group, it was expected that the M group would incur far more heading error since they had to augment their efforts to control the aircraft's roll axis with management of throttle and pitch input as well. These data failed to meet that expectation. The one to three control input ratio was also expected to result in a steeper learning slope for the M group, which also failed to occur as indicated by the lack of any group-trial interaction ($F(5.409, 19)=1.083$, $p>.05$, and the nearly equivalent linear fit slopes indicated in figure 15. Not only do these data support an observed trend towards better learning *without* automation, but the meager effect size associated with them suggest that an increased sample size would have little influence on the results.

Since the basic maneuver scenario for the final training block (TB3) required both groups to execute a descent in full manual mode without turning, all three performance metrics were again evaluated. Although removal of any turning requirement supported the assumption that heading change would be of negligible impact during this task AA(M=0.336, SD=0.797), M(M=0.521, SD=0.802),t(18)=0.516, p>.05, it was expected that the removal of autonomy from the AA group would result in a significant difference in performance across all of the dependent variables in a manner similar to that observed by Clegg & colleagues (Clegg, B.A., et al., 2010). No difference between groups was observed, however, in the number of trials meeting criterion AA(M=1.30, SD=2.163), M(M=5.70, SD=6.093),t(11.23, (18)=0.516, p>.05, or average error incurred across either the altitude AA(M=65.81, SD=37.99), M(M=47.70, SD=24.88),t(18)=1.261, p>.05, or airspeed AA(M=3.949, SD=1.449), M(M=2.847, SD=1.184),t(18)=1.862, p>.05, metrics as indicated in Figure 16.

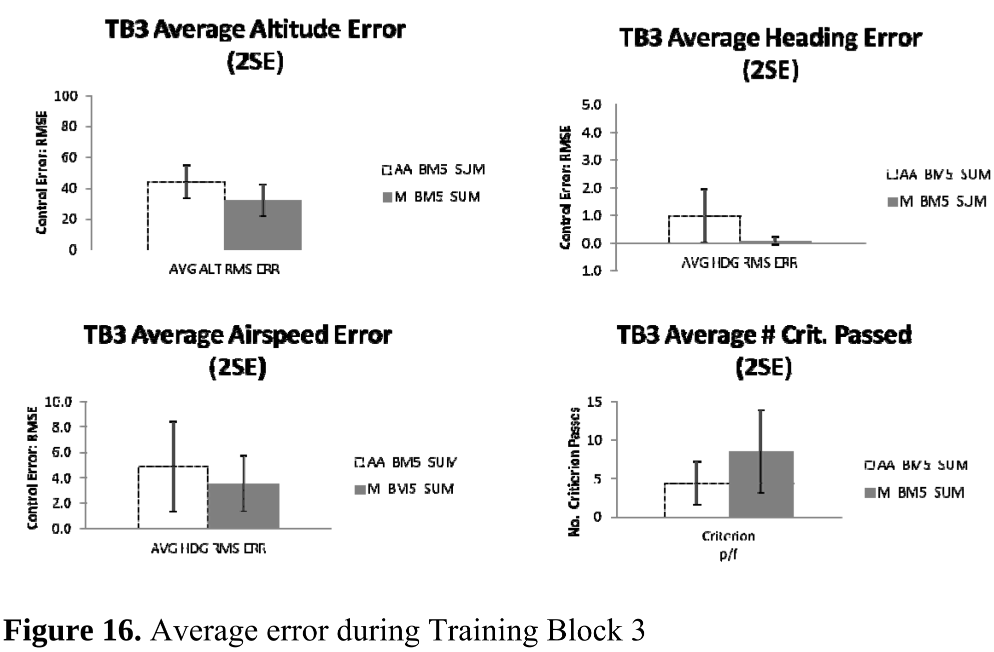

Figure 16. Average error during Training Block 3

A repeated measures ANOVA conducted on a trial by trial analysis also failed to reveal any main effects for group or trial, or any interaction between them for altitude $F(1)= 2.650$, $p>.05$), ($\epsilon = .201$), ($F(3.816, 19)= 1.576$, $p>.05$, ($F(3.816, 19)= 0.906$, $p>.05$ airspeed $F(1)= 0.431$ $p>.05$, ($\epsilon = 0.194$), $F(3.678, 19)=1.970$, $p>.05$, $F(3.678, 19)=0.493$, $p>.05$, or heading $F(1)=3.492$, $p>.05$, $\epsilon = .119$, $F(2.263, 19)= 0.930$, $p>.05$, $F(2.263, 19)= 0.856$, $p>.05$, despite intermittent indications of such in graphic plots of the data shown in figures 17, 18, and 19. As such, these data present no direct evidence of consistent error reduction that would otherwise support the claim that significant learning had taken place during TB3.

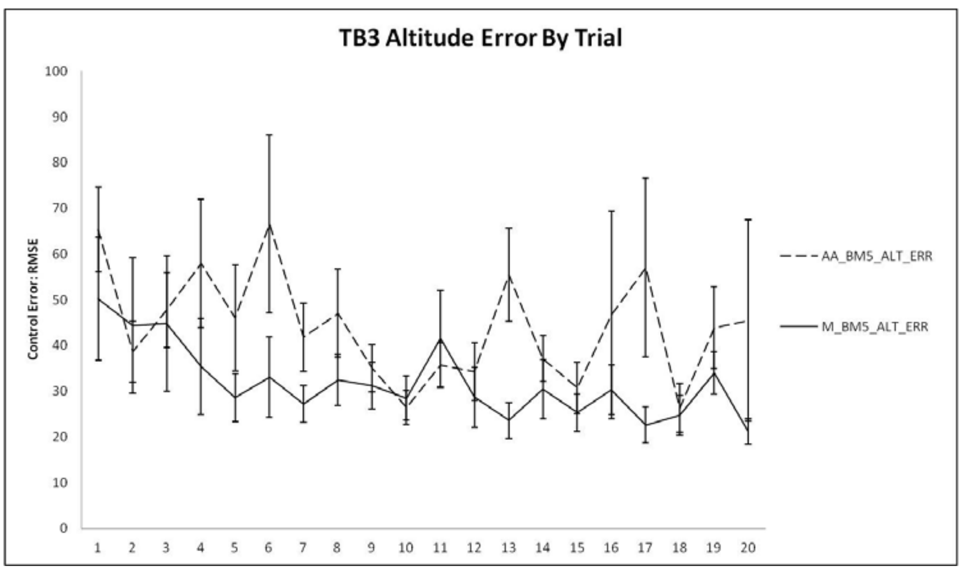

Figure 17. Trial by trial altitude error during Training Block 3

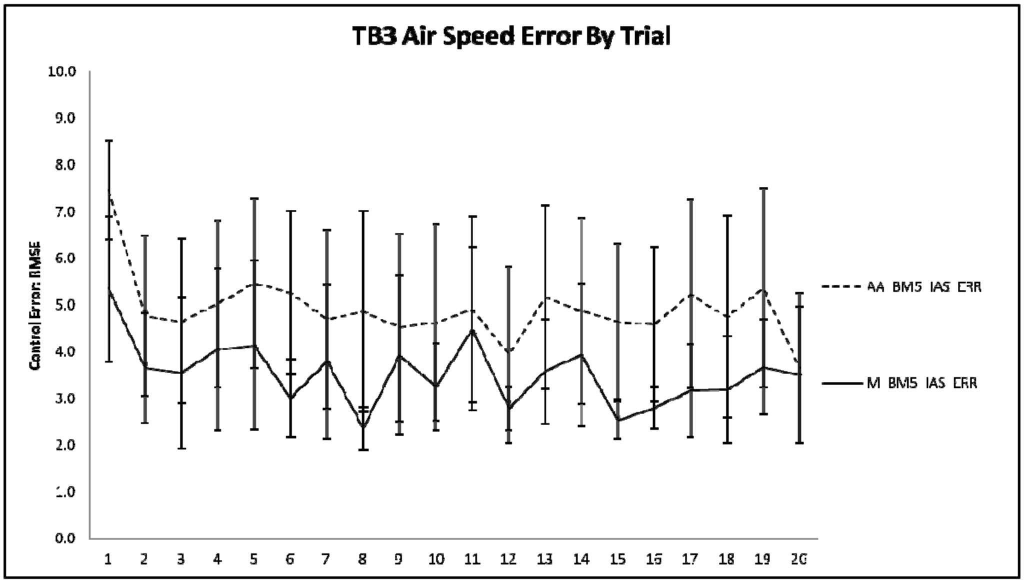

Figure 18. trial by trial airspeed error during training block 3

Even though there was no group difference observed in the average number of times participants were able to reach criterion during their twenty trials, the fact that this occurred approximately thirty percent of the time (M= 6.45, SD=6.94 for both groups combined) provides indirect evidence that learning actually *did* occur during TB3 regardless of whether the trainee had received automation assistance during TB2 or not.

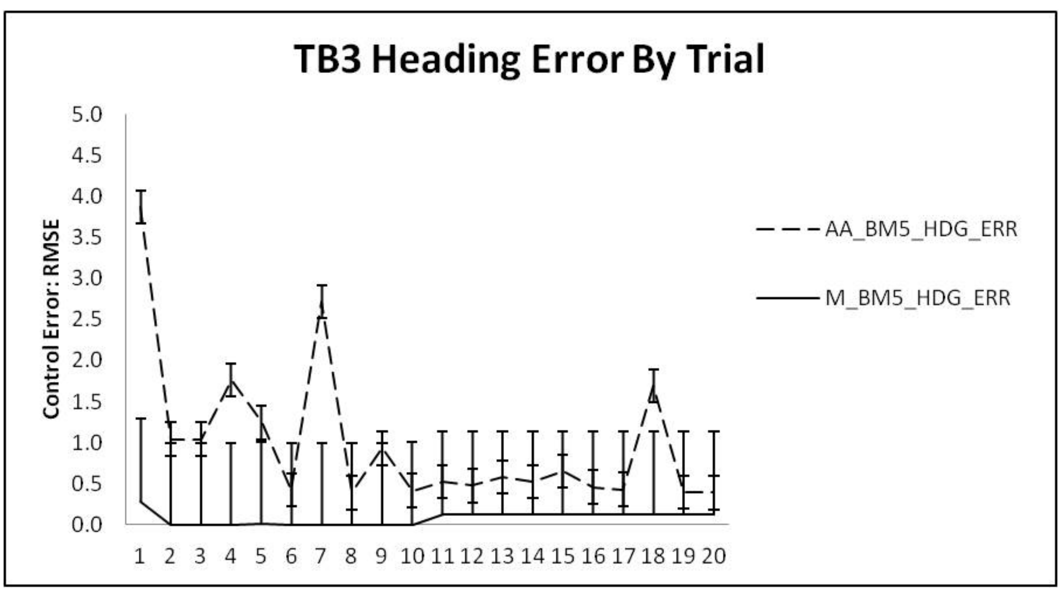

Figure 19. trial by trial heading error during training block 3

Given that criterion achievement is the only performance measure which actually captures how well trainees can *integrate* the three control inputs *simultaneously* by accepting error tradeoffs between individual metrics, it is perhaps the best learning indicator available. It is also worth noting that the 30% criterion observation made in TB3 reflects an appreciable increase over the 15% criterion passed by all participants in TB1, suggesting that learning may be taking place between training blocks as well as within them – particularly since the TB1 and TB3 tasks were of similar difficulty – both requiring a tradeoff between throttle and pitch while holding heading constant.

CHAPTER 4: GENERAL DISCUSSION

Automation influence at test

The magnitude of glide slope error incurred by participants from the automation assisted group compared to those who trained with full manual control presents compelling evidence that learning detriments associated with the automation removal effect reported by Clegg and colleagues (Clegg, Benjamin A., et al., 2010) are not necessarily limited to the domain of process control, and may (based on effect size) degrade training efficiency for unmanned system operators even more severely. Considering the nature of the phrase automation “assistance”, the results from these studies may initially seem a bit counter intuitive. The popular divide and conquer approach to learning complexity, after all, has played a prominent role in learning theory since the advent of cognitive psychology in the 1950s and 1960s (Anderson & Lebiere, 1998, pp. 1-3; Bshouty, 1996; Fu, Lee, Chiang, & Pao, 2001).

From a meta-cognitive standpoint, however, these findings are not all that surprising, particularly in light of more recent literature which has established that humans are notoriously overconfident in many aspects of their learning endeavors (Joseph, 2009; Koriat & Bjork, 2005; Koriat, Bjork, Sheffer, & Bar, 2004; Koriat, Ma'ayan, Sheffer, & Bjork, 2006). Trainees who over estimate their level of capability, after all, may be inclined to remove themselves from voluntary training programs prematurely, thus denying themselves higher performance and retention levels than might otherwise be achieved.

With overconfidence considered in the context of unmanned systems, automation assisted trainees may have only engaged themselves to a minimal extent while maintaining an overly optimistic expectation that they were learning as much as possible. Others may have simply offloaded responsibility for the control error they incurred to imperfect nature of the automation that was supposed to be assisting them in the first place – thereby depriving themselves of sensitivity to subtle yet critical relationships between control inputs that those who struggled with multi-modal control in the manual condition developed a better appreciation for.

When examining the impact of autonomy on training, however, it is important to not only consider the magnitude of its influence on performance, but also nature of how that impact manifests itself in light of manipulated variables. While the performance decrement observed in the autonomy assisted group was expected commensurate with previous findings, it was initially surprising to see the autonomy removal effect achieve statistical significance in only one of the three test performance metrics as opposed to both of the dependent variables measured by Clegg and colleagues – units of good juice produced, and units of spoiled juice produced.

On closer inspection, however, it is interesting to note which performance measures were most influenced by the introduction of automation into the training sequence and the nature of that automation itself. When the autopilot was introduced during TB2, it consisted of functionality applied only to the control inputs (pitch and throttle) which primarily influenced the vertical aspect (or glide slope) of the aircraft's flight path. As such, the autopilot presented a negligible impact on aircraft heading during training which most closely relates to the ground track metrics during test.

The data suggest that by allowing the automation assisted group to focus exclusively on the roll axis during TB2, participants were somehow able to avoid excessive error observed in glide slope, and perform on par with the manual group in both the approach and final ground track metrics. This equivalent proficiency between groups in turning the aircraft, however, appeared to come at a cost in transfer of glide slope control to the landing task test. If automation were not to blame for additional error in the vertical component of flight control, it is doubtful that such a direct mapping of inputs with test data would be evident.

The LTT trial by trial data indicate that the excess error initially incurred by the autonomy assisted group and the beginning of the test phase was reined in a bit over the next four trials. Although it is unclear whether the AA participants were able to completely catch up to the M group in comprehensive performance, the time it took for the AA group to rein in its excess error was comprised of 15-20 minutes of hands on control, or “stick time” spread across four or five LTT trials. This “rein in” period represents approximately 25% - 30% of the 60 x 1 minute basic maneuver training trials – a period which corresponds almost exactly to the length of time that autonomy was involved in TB2. This temporal correlation perhaps lends additional (albeit indirect) support for the link between automation and performance decrements.

The specific reason(s) why this effect persists across both the process control and unmanned aircraft domain at test remains unclear. With an aircraft as complicated to fly as the Predator, one might expect that providing novice trainees with automated assistance would bring an exceptionally challenging performance goal within reach. Instead, it seems to have pushed it further away, at least as far as glide slope management is concerned.

Automation influence during training

While the detrimental impact of autonomy removal appears to be well established at test, its influence during training is less clear. Despite the rigor with which "stick time" was maintained at a constant level, the lack of any significant group differences throughout all three basic maneuver training blocks raises a number of questions about how direct and/or immediate the influence of autonomy is manifested. If, for example, automation influence was tied directly to motivation or various feedback mechanisms, one would logically expect to see an immediate (or nearly so) training performance decrement revealed between groups somewhere in the 40 trials that took place just after autonomy was introduced. Yet the results show no appreciable difference between groups during this period despite concurrent evidence that learning took place during the injection of automation into TB2 and after it was removed.

Automation as a part task training agent

Considered in isolation, the training data do not provide direct evidence of an autonomy removal effect. If participants had become complacent or overly dependent upon autonomy for altitude and airspeed control during TB2, they should have performed significantly worse in TB3 when those responsibilities were suddenly handed back to them. The data simply fail to show such a difference. What is quite clear, however, is that a relative deficiency *did* manifest itself in a closely related component of the landing task as evidenced later by significant differences in glide slope error. One explanation for this perplexing state of affairs can be found in the context of efficiency tradeoffs made between Part Task Training (PTT) decomposition and the integration requirements imposed by Whole Task Training (WTT).

The Part Task Training approach follows a general divide and conquer rubric, handling complex learning challenges by breaking them down into component subtasks that can be practiced in relative isolation. This allows trainees to develop proficiency in a number of basic subtasks before facing the additional (and arguably more complicated) challenge of combining and balancing tradeoffs between them in pursuit of higher goals (Naylor & Briggs, 1963). The Whole Task Training (WTT) paradigm, by comparison, requires trainees to struggle with multiple subtasks simultaneously despite the challenging tradeoffs required between subtasks while simultaneously trying to develop enabling skills themselves (Naylor & Briggs, 1963; Stammers, 1982; Wightman, 1983; Wightman & Lintern, 1985).

Upon closer review of experimental procedures conducted above, it becomes clear that the AA group was never required to integrate all three control inputs in a fashion championed by the Whole Task Training (WTT) realm until they entered the final training/test phase – landing the aircraft. The trial by trial LTT data show that once the AA participants were given the opportunity to attempt this, they were eventually able to rein their excess Glide Slope error when compared to manual participants who were required to integrate all three control inputs (throttle, pitch, & roll) during TB2. In this sense, automation acted as a separation / isolation agent which enabled PTT to be conducted enroute to a WTT objective.

It is important to make a key distinction here between the process control paradigm used by Clegg and colleagues, and PTT/WTT structure used here with the Predator STE. In both cases automation was introduced into the training phase of the experiment and the removed before test. In the process control experiment, however, the evaluation conditions and performance measures remained consistent between the training and test phase. In the composite task used at test in the current experiment, trainees were not only faced with a new scenario that required component skills to be balanced and integrated in a novel (though arguably familiar) way, but their performance was evaluated by different metrics as well.

This subtle yet significant difference in experimental design may provide some insight as to why the autonomy removal effect was so prominent in the process control experiment's training phase but not here. The trainees in the process control arrangement were oriented on an aggregate task that may have induced a relative dependency on autonomy which caused an obvious and immediate drop in training performance when suddenly removed. The Predator STE experiment, by comparison, was oriented on a highly organized integration task at test which required trainees to not only combine their newly acquired skills, but balance them with overlapping performance tradeoff considerations developed during their previous three training blocks.

The automation assisted group was spared the requirement to make such tradeoffs during TB2 creating a situation which presented no observable decrement in performance at the time, but appeared to result in an impoverished mental model of the relationship between component skills compared with the manual group. This created a performance deficit at test that was surprisingly strong and persistent across multiple trials. The durability of this effect essentially kicked the integration can down the road at an alarming cost in overall training efficiency. While the participants exposed to automation were eventually able to reduce their error rate to one roughly approximating that of the manual group, it took them 4 additional trials to do so – a cost that any evaluator or training manager would be reluctant to accept.

This delayed onset of this performance degradation also raises concerns regarding the potential for automation to mask deep deficits in conceptual understanding during component training activities. The sudden appearance of these deficits during more complex endeavors with higher task organization is of particular concern for trainers trying to prepare operators for future duties in high risk environments where even moderate performance decrements can result in the catastrophic loss of innocent lives (Massood & Shah, 2011; Mazetti & Schmitt, 2011; Smith, 2011).

Automation and workload

Given the part task / whole task implications discussed above, the question remains whether underlying principles can be identified for autonomy induced learning deficits. Previous work suggests that the injection of an automated module into our training regimen may have pulled the AA group out of a desirable difficulty sweet spot that is required to develop comprehensive mental models via deep encoding during learning (Bainbridge, 1983; Endsley & Kiris, 1995; Moray, 1986).

The notion of desirable difficulty has been considered an important aspect of training research and skill acquisition at least as far back as 1956, when it was discovered that more skill was often transferred between tasks when a more difficult version was presented first rather than the other way around (Day, 1956). The positive influence of difficulty was also noted in the recall accuracy of material presented before and after an intervening task was injected into a typical study / test process. Participants who were required to perform a difficult task in between material presentations recalled more information at test than those afforded an easier intervening task (Bjork & Allen, 1970; Carpenter & DeLosh, 2006; Schmidt & Bjork, 1992).

A general consensus in the literature considers difficult yet attainable tasks to often yield the most efficient learning (Locke, Shaw, Saari, & Latham, 1981), albeit subject to influence and compatibility with established goals (Huber, 1985). Consistent improvements in test performance and delayed retention have also been associated with the introduction of difficulty during training in both motor and verbal task domains *even when doing so appears to lower performance during training* (Bjork & Bjork, 2006; Schmidt & Bjork, 1992).

Considered in the context of this research foundation and the framework presented here, the results suggest that the introduction of automation imposed a performance deficit at test by 1) making the training during TB2 too easy for even novice participants to develop a comprehensive mental model of aircraft response to various control input combinations, and by 2) denying them the experience of managing those inputs under challenging conditions that would allow them to perform as well as the manual group during landing.

Alternative explanations

Motivational issues are often postulated as a cause and/or consequence of automation induced complacency, especially in monitoring or process control tasks, which may drive systems design toward redesign of feedback as an appropriate mitigation strategy (Moray, 1986; Norman, D. A., 1990). It has been suggested, for example, that a trainee's perceived shift in locus of control (to the autopilot in this case) may influence their motivation to learn, especially in regard to feedback (Noe, 1986). Since trainees in this experiment received feedback on each and every trial they performed, these issues are particularly important to consider as influential factors.

Given the 75 times that participants received end of trial performance feedback, for example, it is quite possible that those who perceived an improvement in their performance might have become apathetic when unrequested (and perhaps even unwanted) “assistance” from the autopilot was thrust upon them in TB2. Given that they were specifically told to ignore trial by trial feedback on the two performance components that the autopilot was supposed to perform for them (altitude and airspeed hold) this indifference prediction seems likely. If particularly perceptive participants ignored instructions and monitored the autopilot’s performance anyway, they might have even begun to realize that the autopilot actually performed these functions in an imperfect fashion. Upon realizing that the autopilot’s assistance might prevent them from ever reaching criterion during this training block (which was the case, but participants were not explicitly told this) the AA group would have been understandably justified in adopting an apathetic attitude and simply giving up on the task.

Similar arguments have been made in discussion of automation induced complacency and other influential factors that may reduce situational awareness of the aircraft operator, creating a dreaded “out of the loop” performance decrement in the process (Endsley & Kiris, 1995; Kaber, et al., 2000). Since we did not explicitly measure personal attitude, motivation, or situational awareness we had no direct empirical way to test this apathy or “out of the loop hypotheses. Indirect analysis of the performance data that was collected, however, does not support it.

The fact that there no group performance differences were observed after automation was introduced in TB2 or after it was removed in TB3 drains any support for the notion that a significant change in motivation occurred. Since the number of trials in which participants reached criterion in TB3 actually *increased* compared to previous training blocks, it is highly unlikely that motivation to learn decreased after automation was introduced or subsequently removed. These results also suggest that the automation assisted participants remained aware enough of their situation to keep up with the learning rate achieved by the manual group. It follows from the increase in trials reaching criterion for both groups during TB3 that motivation and awareness remained intact or perhaps even increased slightly after automation was invoked. Anecdotal evidence for this was provided via research logs on which experimenters were required to record any apparent signs of duress, distraction, or motivational deficits during each trial they observed. No behavioral indicators of motivation deficits were noted for either group.

With regard to relationships that may exist between the autopilot's influence on motivational factors and locus of control, it might have been interesting to examine whether user invoked automation could result in different learning patterns than full manual control or automatically applied automation as Clegg and colleagues did in the orange juice pasteurizer study. The inherent software limitations of the Predator STE preclude this type of experimental design, however, because the auto pilot has to be setup prior to the beginning of each trial and cannot be adjusted or even turned off until the trial ends. Thus, the participants in this experiment had no capacity to toggle the autopilot on or off once a trial had started, or even change which control inputs they would be allowed to manage, so the influence of variable activation remains unexplored here and left to future research with different simulators.

There are attentional switching and focus issues to be explored here however, particularly as they pertain to the two computer monitors which participants used for awareness of the aircraft's status as indicated in Figure1. Such concerns are of prominent relevance under conditions with multiple task loads and variable control requirements that are often competing for attentional resources (Parasuraman & Manzey, 2010; Wickens, et al., 2006).

During TB1, the left monitor contained all the instrumentation needed to hold the aircraft's heading and altitude steady while reducing airspeed from 67 down to 62 knots. This instrumentation consisted of the artificial horizon, current airspeed, attitude indicator, vertical speed indicator, and current altitude display as indicated by the red dashed arrows in Figure 20. Since heading was supposed to be held constant there was no information displayed on the right monitor that contributed directly to this task. During training it was often observed that participants would focus almost exclusively on the left monitor for the entire trial, switching their focus to the right monitor only at the end to receive feedback.

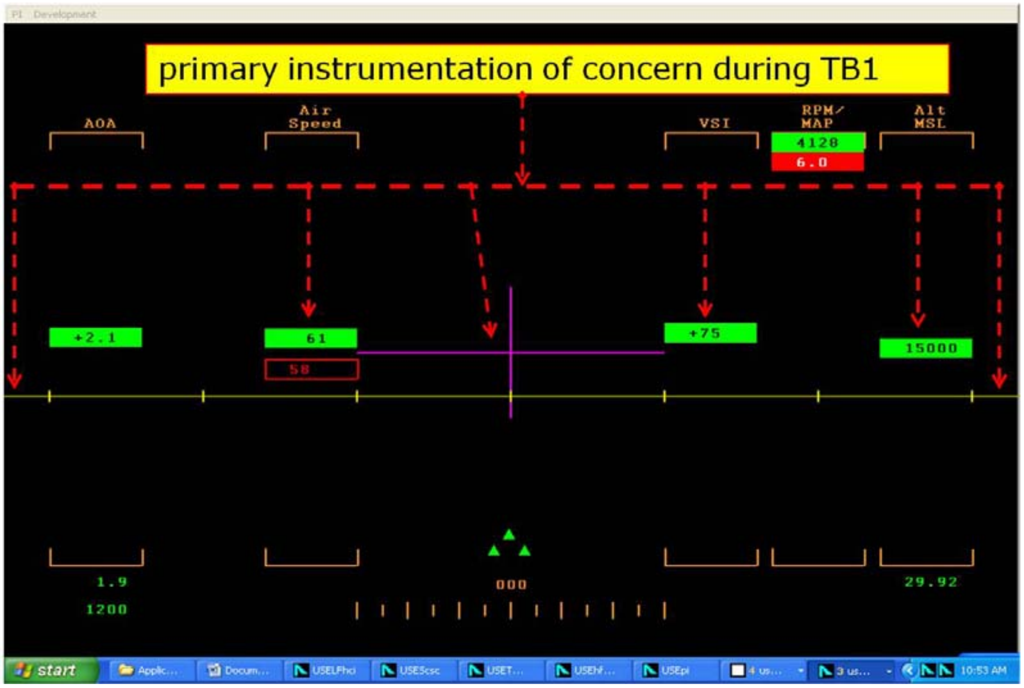

Figure 20: Typical left monitor display at the beginning of TB1 trials

Because performance data was always displayed on the right monitor at the completion of each training or landing trial, it is quite possible (if not probable) that participants became conditioned during TB1 to focus on the left monitor during simulated flight and only use the right monitor at the end of each trial for feedback. The turning requirement initiated in TB2, however, presented an advantage of switching attention to the right monitor from time to time. As participants transitioned into this turning scenario, they needed to expand their attentional resources to include additional instrumentation displayed on the monitor in the form of the numerical heading and rate of turn indicators highlighted by red circles in Figure 21. Although the left monitor provided trainees with sufficient information to create a gross representation of the aircraft's posture during a turn, the right monitor presented additional information with the potential for a refined posture to be ascertained in pursuit of the critical error reduction necessary to reach criterion.

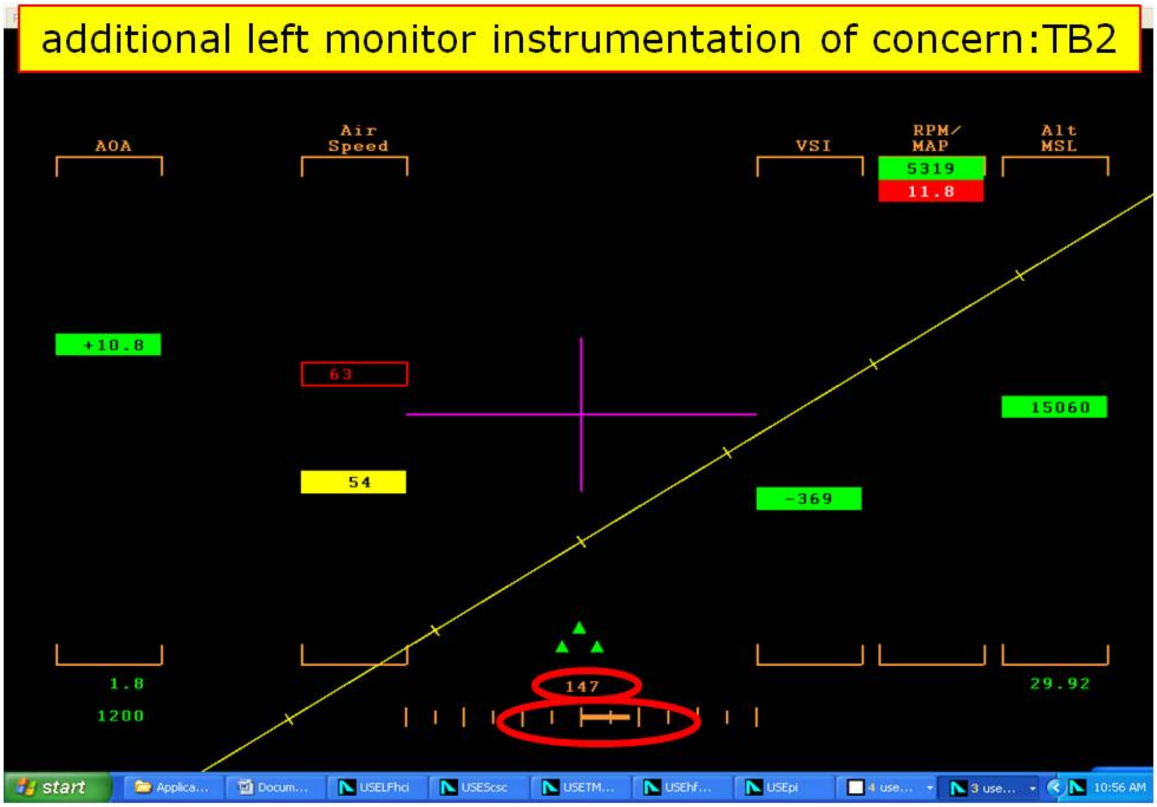

Figure 21: Typical left monitor display during TB2 trials

On the left monitor heading is represented numerically while bank angle and turn rate are represented graphically in the form of the tilted yellow artificial horizon and thick orange slider bar at the bottom of the screen. The right monitor reverses this arrangement by adding a numerical component for bank angle and a compass ring for heading as indicated in Figure 22. What is so important about this aspect of turn co-ordination is the relationship between bank angle and turn rate. Trainees typically struggle to establish and maintain the proper turn rate during this task as evidenced by substantial oscillations in bank angle, especially when having to control for altitude and airspeed drift at the same time. Only those trainees who are able to rapidly but smoothly establish and maintain the standard rate of turn (3 degrees per second) throughout the majority of the trial are able to reach criterion in this task.

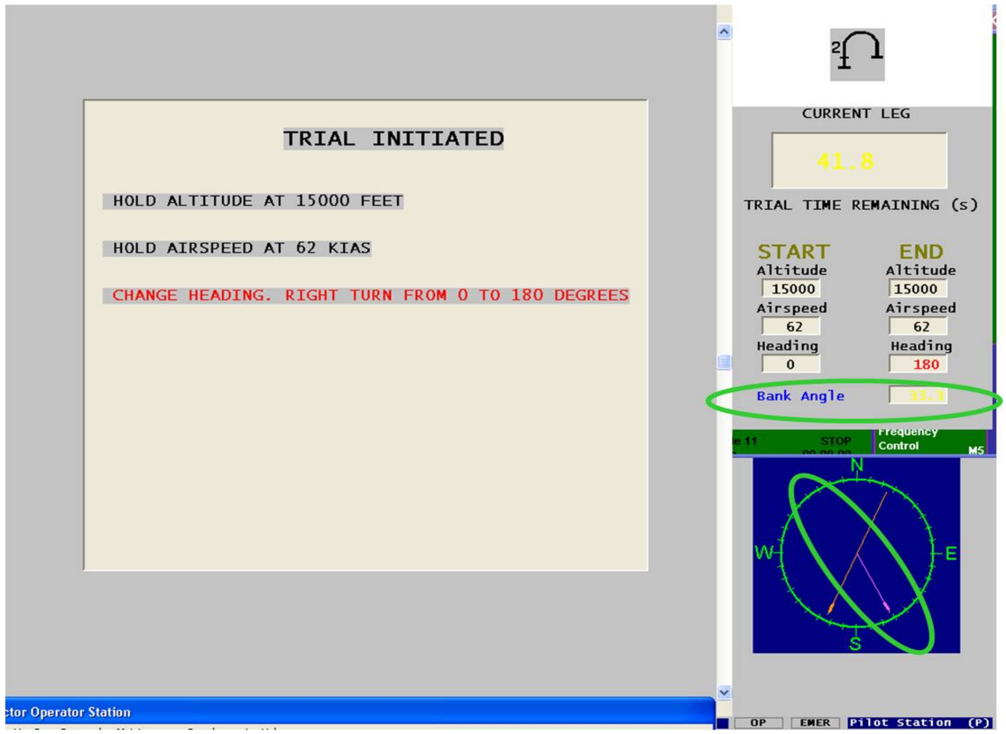

Figure 22: Typical right monitor display during TB2 trials

Since there is no numerical representation of turn rate anywhere in STE's display, trainees are forced to estimate it graphically via extension of the thick yellow bar at the bottom of the screen – which provides only a coarse estimation at best. The next best input for turn rate estimation is bank angle, but that representation on the left monitor is presented in a coarse analog graphic as well. The right monitor, however, presents bank angle in a crisp numerical fashion that requires little interpretation. It also presents a macro view of heading in a natural graphic that most people are familiar with regardless of their aviation experience – the compass rose.

It has been well established in the perceptual and attention literature that increases in task difficulty typically result in a more narrow perceptual focus, while decreased difficulty is associated with greater generality of transfer between tasks (Ahissar & Hochstein, 1997, 2000; Wickens & Andre, 1990). Given this background, it is doubtful that the manual group was inclined to glance over to the right monitor for additional information while struggling to control pitch, throttle and bank angle control inputs while simultaneously monitoring all of the other display items available during TB2.

The automation assisted group, on the other hand, only needed to monitor those display features directly related to the single axis of control (roll/bank angle) for which they were responsible. The freedom from multi-task control could have provided the AA group with the opportunity to view the additional information available on the right monitor and thus enjoy a corresponding reduction in their heading error during TB2. This is precisely the kind of automation induced mechanism that has been shown to actually *increase* situational awareness in the performance of complex tasks (Endsley & Kaber, 1999; Wickens, 2008). It also suggests that some participants in AA group may actually have been learning different cues and indirectly practicing different task variations than the manual group under Sheridan's notion of automation influence (Sheridan & Parasuraman, 2000; Sheridan & Parasuraman, 2005).

The problem with these hypotheses, however, is that the data simply fail to support them. If the AA group had suffered an out of the loop effect, then their heading error during TB2 should have been significantly larger than that found for the manual group. If they had been able to capitalize on the reduced demand for their attentional resources and exploit the additional information on the right monitor, then they should have performed significantly better than the manual group in TB2. The fact that no performance differences were found between groups during TB2 or TB3 denies support for either of these situations. The lack of attentional / situational awareness measurements via eye tracking or other devices also precludes the development of any direct empirical determination regarding the bearing of these factors on the results obtained, and serves as additional motivation for future research.

Conclusion and implications

Taken collectively, the results reported above support the hypothesis that autonomy removal effects observed in the process control domain do in fact transfer to the motor control arena exemplified by the Predator STE unmanned systems simulator. They also exposed a troubling feature of automation induced training deficiency in that its deleterious influence was apparently masked during training itself. The most parsimonious account of this finding resides in the notion that automation served as a part task training agent which interfered with the development of a comprehensive mental model during learning that did not manifest itself until the integration of component skills was required at test. Additional support for this interpretation comes from a tank gunnery study in which automation was invoked in a similar manner to moderate part/whole training effects (Marmie & Healy, 1995), albeit in an extended retention format that lies beyond the scope of this experiment.

A theoretical basis for the poor mental model formulation evident at test, but not during training, may be examined in terms of the distinction between intrinsic, germane, and extraneous cognitive load offered by proponents of Cognitive Load Theory (CLT). This theory, fashioned loosely on a brain function analogy that emulates a generator style load management paradigm, presents intrinsic load as a basic component of learning which differs significantly depending upon the interactivity between various knowledge elements (Paas, 1992). Tasks with low interactivity between elements tend to be learned serially in a rote memory fashion, while those with high interactivity can actually be understood at a high level, and recorded in the form of various schemas which can be processed with varying degrees of automaticity by the learner (Sweller, van Merrienboer, & Paas, 1998). Schema construction, which lies at the heart of the theory, benefits most from germane load in which understanding of interactivity between elements is maximized, and least from extraneous load which generally serves to divert the learner's attention away from it.

This appreciation for interactivity between training elements fits well with the discussion of sensitivity to control mode interaction presented above. It follows from that discussion that automation somehow interfered with the understanding that should have been derived from direct interactivity experience between control mode "elements" in the manual group. This in turn may have deprived the automation group of sufficiently well developed schemas that they would need to transfer into a new conceptual model during the novel landing task presented at test. Unfortunately, the innovative and logical scaffold beneath CLT is still lacking empirical support from dissociations between and direct measures of cognitive load variants during training (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). It is hoped that the findings presented here may provide a modicum of such support in future evolution of the theory.

These findings also introduce some important implications for trainers of the future – especially those hoping to increase efficiency by providing automated assistance to novice trainees in the form of complex task decomposition. Not only was evidence of any automation advantage absent during the entire four hour procedure followed here, but the substantially detrimental effect observed in the opposite direction was not revealed until the very last test phase of training. This not only provides evidence that automation can indeed mask learning deficits in the unmanned systems control regime on par with those previously observed in the process control and decision support domains (Bainbridge, 1983; Moray, 1986) – it begs the question as to whether similar effects may remain hidden in other learning endeavors as well.

Perhaps the most dangerous aspect of all, however, is the implication that some trainees tend to offload responsibility for the error they incur as an artifact of imperfect automation, even when it is activated at a minimal level as observed here during the 20 TB2 trials. Such an effect could transfer detrimental influences beyond component skill acquisition into higher forms of integrated reasoning, such as those required for ethical and moral decision making. The appalling nature of people's obedience to authority observed in Stanley Milgram's infamous shock experiments, for example, is widely considered to represent a human tendency to offload responsibility for undesirable consequence to others (Blass, 2004). Perhaps this tendency would be evident (if not amplified) in unmanned system operators involved in situations where autonomous agents of some sort were "wearing" the lab coat instead of humans. Despite Hollywood hype that surrounds various movies projecting robots and automated systems that have somehow evolved to acquire lethal capabilities, additional caution may be in order regarding just how much of accountability we allow ourselves to offload to machines. With a heavy caveat that these studies only present undergraduate student data from a single university, these findings support a warning flag with a least a modicum of visibility be waved in military circles.

Future work: toward a neuro-adaptive training workload sweet spot

Despite the negative connotation of the findings described above, the tremendous potential for automation to improve both the efficiency and effectiveness of learning is indisputable. From the very advent of computer science, Ed Feigenbaum proved that an expert system comprised of rule base extracted from multiple doctors' diagnostic experience could consistently outperform junior physicians and occasionally even those with the most impressive track record and profession background (Russell, Norvig, & Artificial Intelligence, 1995, pp. 23-24). For those learning how to diagnose symptoms associate complex problems, this notion of encapsulated experience cannot escape consideration as a powerful pedagogical tool. Even those exposing serious concerns for automation's impact on transportation safety cannot help but provide a balanced perspective on its value (Endsley & Kaber, 1999; Parasuraman & Wickens, 2008). From an applied tactical/rescue research perspective, automation is not only an important factor for unmanned systems design – it represents a *critical* component of its very essence as a reliable extension of human capability.

Some in the human factors realm have defined automation as "the execution by a machine agent (usually a computer) of a function that was previously carried out by a human (Parasuraman & Riley, 1997, p. 231). Blitch, however, rejected this limited, anthro-centric perspective in a vigorous plea for leaders within the military, emergency services, and space exploration communities to embrace the powerful potential of mechanical design freedom in pursuit of advanced robotic capabilities that could not only replicate human performance, but actually *exceed* it (Blitch & Maurer, 1996; Krotkov & Blitch, 1999; Weisbin et al., 1999). This work originally championed the deployment of "micro" robots (loosely defined as having a cross section well below a human form factor) for penetration of denied areas on the battlefield, in narrow space habitat conduits and rock fissures potentially hiding UV-phobic microbes on Mars, and twisted void spaces in collapsed rubble. Later efforts, however, focused on development of Adaptive Robotic Manipulators (A.R.Ms) that might combine the amazing extension capacity of cephalopod tentacles and other muscular hydrostats with the hyper redundant nature of elephant trunks to create a family of robot appendages with a wrapping capability for compliant object manipulation, dynamic target capture, and other tasks of interest to those working in these challenging operational domains. Unlike the cockpit derivations of Fitts' list that focused on the liberation of cognitive resources from tedium in pursuit of greater efficiency and human *convenience*, the evolution of imperatives for Tactical Mobile Robot (TMR) control anticipated automation as a *critical* technology which actually *enabled* control of complicated systems such as a multi-limbed collection of squid tentacles that evolutionary models of human cognition were incapable of (Blitch, 2000).

Considering such "inhuman factors" in a training environment where robot operators must learn how to control hyper-redundant, multi-limbed manipulators that exceed human capacities derived from anthropomorphic evolution, the injection of automation into training is unavoidable. This inevitability and the alternative hypotheses discussed heretofore dictate the need for continued research into identification and mitigation strategies for whatever deleterious influence automation may occasionally exert on trainees in this challenging and important arena. In addition to the pursuit of higher statistical power, extensive workload measurement must be conducted enroute to a more comprehensive understanding of the underlying principles involved. Given the well established tradeoff between immediate skill acquisition and retention durability (Healy, Ericsson, & Bourne Jr, 1999; Healy, Wohldmann, Parker, & Bourne, 2005), retention of the skills acquired here should also be assessed over various duration intervals.

Returning to the notion of desirable difficulty and cognitive load, it seems that the most compelling issue to address at this point is how to promote the formulation of accurate schemas integrate them into comprehensive cognitive models while avoiding automation induced pitfalls enroute to more skill and expertise. Although Ericsson's seminal work on expertise acquisition is often cited for its numerical threshold concerning ten thousand hours of practice, few recognize the emphasis placed on deliberate focus attached to that practice (Ericsson, Krampe, & Tesch-Römer, 1993). Fewer still demonstrate an appreciation for the nature of what sort of practice can be considered "deliberate" in the first place (Gladwell, 2008). In pursuit of the sweet spot of desirable difficulty described early, it seems logical that automation introduced to trainees for pedagogical purposes will need to adapt to the student's workload level in order to pursue optimal learning(Van Merriënboer, Kester, & Paas, 2006).

The need for adaptive automation has been well established and examined across a number of environmental settings, including the process control (Moray, Inagaki, & Itoh, 2000) and aviation (Parasuraman, 1992) domains which were of particular concern here. Literature regarding adaptive automation involved in novice training activity is scarce, however, and often deals more with resource allocation activities than direct control of unmanned platforms (Kaber & Endsley, 2004).

Given that automation induced learning deficits that are masked during training can manifest in potentially dangerous ways downstream, adaptive modules for enhanced learning will most likely need to monitor inputs other than those based on performance in order to promote efficient schema formulation without negative side effects. Recent development of neuro-imaging techniques suitable for monitoring workload in addition to task performance can provide an appreciable degree of optimism in addressing this challenging objective (Berka et al., 2004; Berka et al., 2007; Freeman, Mikulka, Prinzel, & Scerbo, 1999). In any case, there is ample evidence here for caution to be applied when automation is injected into the learning environment. In a world where the mystical nature of artificial intelligence has enticed some to turn over classrooms to robot teachers (Agostini, Celaya Llover, Torras, & Wörgötter, 2008; Carey & Markoff, 2010), the stakes could not be higher for our children's children.

REFERENCES

Agostini, A. G., Celaya Llover, E., Torras, C., & Wörgötter, F. (2008). Action rule induction from cause-effect pairs learned through robot-teacher interaction. *Proceedings of*.

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature, 387*, 401-406.

Ahissar, M., & Hochstein, S. (2000). The spread of attention and learning in feature search: effects of target distribution and task difficulty. *Vision Research, 40*(10-12), 1349-1364.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*: Lawrence Erlbaum.

Antal, J. (2009). I Fight the Body Electric! [Article]. *Military Technology, 33*(7), 22-30.

Bailey, N. R., & Scerbo, M. W. (2007). Automation-induced complacency for monitoring highly reliable systems: the role of task complexity, system experience, and operator trust. [Article]. *Theoretical Issues in Ergonomics Science, 8*(4), 321-348. doi: 10.1080/14639220500535301

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*(6), 775-779. doi: Doi: 10.1016/0005-1098(83)90046-8

Baker, M., Casey, R., Keyes, B., & Yanco, H. A. (2004). Improved interfaces for human-robot interaction in urban search and rescue. *Proceedings of*, 2960-2965 vol. 2963.

Ballas, J. A., Heitmeyer, C. L., & Pérez-Quiñones, M. A. (1992). Evaluating two aspects of direct manipulation in advanced cockpits. *Proceedings of*, 127-134.

Barab, S. A., Scott, B., Siyahhan, S., Goldstone, R., Ingram-Goble, A., Zuiker, S. J., et al. (2009). Transformational play as a curricular scaffold: Using videogames to support science education. *Journal of Science Education and Technology, 18*(4), 305-320.

Berka, C., Levendowski, D., Cvetinovic, M., Petrovic, M., Davis, G., Lumicao, M., et al. (2004). Real-Time Analysis of EEG Indexes of Alertness, Cognition, and Memory Acquired With a Wireless EEG Headset. *International Journal of Human-Computer Interaction, 17*(2), 151-170.

Berka, C., Levendowski, D., Lumicao, M., Yau, A., Davis, G., Zivkovic, V., et al. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine, 78*(Supplement 1), B231-B244.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. M. A. Shimamura (Ed.), *Metacognition: Knowing About Knowing* (pp. 185-205). Cambridge MA: The MIT Press.

Bjork, R. A., & Allen, T. W. (1970). The spacing effect: Consolidation or differential encoding. *Journal of Verbal Learning & Verbal Behavior, 9*(5), 567-572. doi: 10.1016/s0022-5371(70)80103-7

Bjork, R. A., & Bjork, E. L. (2006). Optimizing treatment and instruction: Implications of a new theory of disuse. In L. G. Nilsson & N. Ohta (Eds.), *Memory and society: Psychological perspectives.* (pp. 116-140). New York, NY US: Psychology Press.

Blass, T. (2004). *The man who shocked the world: The life and legacy of Stanley Milgram*: Basic Books New York.

Blitch, J. G. (2000). Adaptive Robotic Manipulators *DARPA Workshop On Advanced Manipulation*. Marine Biological Laboratory (MBL), Woods Hole MA: Defense Advanced Research Projects Agency.

Blitch, J. G., & Maurer, R. (1996). KNOBSAR: a knowledge based system prototype for robot assisted urban search and rescue. [Feature]. *Simulation, 66*, 375-391.

Broadhurst, P. (1959). The interaction of task difficulty and motivation: The Yerkes-Dodson Law revived. *Acta Psychologica, 16*, 321-338.

Bshouty, N. H. (1996). Simple learning algorithms using divide and conquer. *Computational Complexity, 6*(2), 174-194.

Canning, J. (2008). Weaponized Unmanned Systems: A Transformational Warfighting Opportunity, Government Roles in Making it Happen. *preprint to appear in American Society of Naval Engineers Journal*.

Carey, B., & Markoff, J. (2010). Students, Meet Your New Teacher, Mr. Robot, *New York Times*. Retrieved from www.nytimes.com/2010/07/11/science/11robots.html

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34*(2), 268-276.

Clegg, B. A., & Heggestad, E. D. (2010). Automation and effective training: Colorado State University Technical Report for U.S. Army Research Office MURI Grant W911NF-05-1-0153.

Clegg, B. A., Heggestad, E. D., & Blalock, L. D. (2010). The Influences of Automation and Trainee Aptitude on Training Effectiveness. *Journal of the Human Factors and Egonomics Society*.

Clegg, B. A., Heggestad, E. D., & Blalock, L. D. (2010). The Influences of Automation and Trainee Aptitude on Training Effectiveness. *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting, 54*(27), 2329-2332.

Crawford, J. M., & Weisbin, C. R. (2005). Autonomous Operations for the Crew Exploration Vehicle — Trade Study Design Considerations. [Article]. *AIP Conference Proceedings, 746*(1), 1145-1152. doi: 10.1063/1.1867241

Day, R. (1956). Relative task difficulty and transfer of training in skilled performance. *Psychological Bulletin, 53*(2), 160-168.

Endsley, M. (1997, June 1997). Situation awareness, automation and free-flight*. Proceedings of FAA/Eurocontrol Management R&D Seminar*, Saclay, France.

Endsley, M., & Kaber, D. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics, 42*(3), 462-492.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors, 37*(2), 381-394. doi: 10.1518/001872095779064555

Ericsson, K., Krampe, R., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *PSYCHOLOGICAL REVIEW-NEW YORK-, 100*, 363-363.

Freeman, F. G., Mikulka, P. J., Prinzel, L. J., & Scerbo, M. W. (1999). Evaluation of an adaptive automation system using three EEG indices with a visual tracking task. *Biological Psychology, 50*(1), 61-76.

Fu, H. C., Lee, Y. P., Chiang, C. C., & Pao, H. T. (2001). Divide-and-conquer learning and modular perceptron networks. *Neural Networks, IEEE Transactions on, 12*(2), 250-263.

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE), 1*(1), 20-20.

Gladwell, M. (2008). *Outliers: The story of success*: Little, Brown and Company.

Gramm, J., & Papp, S. (2009). *An Insatiable Demand:'Manning'the US Air Force's Unmanned Aircraft Systems with Capable Pilots.* Masters Degree in Public Policy, Harvard Kennedy School of Government

Haider, H., & Frensch, P. A. (2002). Why aggregated learning follows the power law of practice when individual learning does not: Comment on Rickard (1997, 1999), Delaney et al. (1998), and Palmieri (1999). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(2), 392-406. doi: 10.1037/0278-7393.28.2.392

Healy, A., Ericsson, K., & Bourne Jr, L. (1999). Optimizing the Long-Term Retention of Skills: Structural and Analytic Approaches to Skill Maintenance. Annual Report, 1991-1992.

Healy, A., Wohldmann, E., Parker, J., & Bourne, L. (2005). Skill training, retention, and transfer: The effects of a concurrent secondary task. *Memory & Cognition, 33*(8), 1457.

Healy, A. F., Buck-Gengler, C. J., Barshi, I., Parker, J. T., Schneider, V. I., Raymond, W. D., et al. (2002). Optimizing the durability and generalizability of knowledge and skills. In S. P. Shohov (Ed.), *Trends in cognitive psychology.* (pp. 123-192). Hauppauge, NY US: Nova Science Publishers.

Healy, A. F., Kole, J. A., Wohldmann, E. L., Buck-Gengler, C. J., Parker, J. T., & Bourne Jr, L. E. (2005). Optimizing the speed, durability, and transferability of training (pp. 135): Lawrence Erlbaum.

Healy, A. F., Wohldmann, E. L., & Bourne, L. E., Jr. (2005). The Procedural Reinstatement Principle: Studies on Training, Retention, and Transfer. In A. F. Healy (Ed.), *Experimental cognitive psychology and its applications.* (pp. 59-71). Washington, DC US: American Psychological Association.

Hennigan, W. J. (2012, 26 January 2012). New drone has no pilot anywhere, so who's accountable?, *Los Angeles Times*.

Huber, V. (1985). Effects of task difficulty, goal setting, and strategy on performance of a heuristic task. *Journal of Applied Psychology, 70*(3), 492-504.

Joseph, N. (2009). Metacognition Needed: Teaching Middle and High School Students to Develop Strategic Learning Skills. [Article]. *Preventing School Failure, 54*(2), 99-103.

Kaber, D., Onal, E., & Endsley, M. (1999). Level of automation effects on telerobot performance and human operator situation awareness and subjective workload. *Automation technology and human performance: Current research and trends*, 165-169.

Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science, 5*(2), 113-153.

Kaber, D. B., Onal, E., & Endsley, M. R. (2000). Design of automation for telerobots and the effect on performance, operator situation awareness, and subjective workload. *Human Factors and Ergonomics in Manufacturing, 10*(4), 409-430.

Koriat, A., & Bjork, R. A. (2005). Illusions of Competence in Monitoring One's Knowledge During Study. [Article]. *Journal of Experimental Psychology / Learning, Memory & Cognition, 31*(2), 187-194. doi: 10.1037/0278-7393.31.2.187

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting One's Own Forgetting: The Role of Experience-Based and Theory-Based Processes. [Article]. *Journal of Experimental Psychology / General, 133*(4), 643-656. doi: 10.1037/0096-3445.133.4.643

Koriat, A., Ma'ayan, H., Sheffer, L., & Bjork, R. A. (2006). Exploring a Mnemonic Debiasing Account of the Underconfidence-With-Practice Effect. [Article]. *Journal of Experimental Psychology / Learning, Memory & Cognition, 32*(3), 595-608.

Kornell, N., & Bjork, R. A. (2008). Optimising self-regulated study: The benefits—and costs—of dropping flashcards. *Memory, 16*(2), 125-136.

Krotkov, E., & Blitch, J. (1999). The Tactical Mobile Robotics Program. *The International Journal of Robotics Research, 18*(7), 769-776.

Kurzweil, R. (2005). *The singularity is near*: Penguin books.

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics, 35*(10), 1243-1270.

Lee, J. D. (2006). Human factors and ergonomics in automation design. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics (3rd ed.).* (pp. 1570-1596). Hoboken, NJ US: John Wiley & Sons Inc.

Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969–1980. *Psychological Bulletin, 90*(1), 125-152. doi: 10.1037/0033-2909.90.1.125

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human-human and human-automation trust: An integrative review. *Theoretical Issues in Ergonomics Science, 8*(4), 277-301. doi: 10.1080/14639220500337708

Martin, E., Lyon, D., Schreiber, B., & Martin, L. (1998). *Designing synthetic tasks for human factors research: An application to uninhabited air vehicles*. Paper presented at the Human Factors and Ergonomics Society Annual Meeting.

Massood, S., & Shah, P. (2011, March 17 2011). C.I.A. Drones Kill Civilians in Pakistan, *New York Times*. Retrieved from http://www.nytimes.com/2011/03/18/world/asia/18pakistan.html

Mazetti, M., & Schmitt, E. (2011, 02 July 2011). U.S. Expands Its Drone War Into Somalia, *New York Times*. Retrieved from http://www.nytimes.com/2011/07/02/world/africa/02somalia.htm

Micire, M., Drury, J. L., Keyes, B., & Yanco, H. A. (2009). Multi-touch interaction for robot control. *Proceedings of*, 425-428.

Moray, N. (1986). Monitoring behavior and supervisory control. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of perception and human performance, Vol. 2: Cognitive processes and performance.* (pp. 1-51). Oxford England: John Wiley & Sons.

Moray, N., & Inagaki, T. (1999). Laboratory studies of trust between humans and machines in automated systems. *Transactions of the Institute of Measurement and Control, 21*(4-5), 203.

Moray, N., Inagaki, T., & Itoh, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology: Applied, 6*(1), 44.

Naylor, J. C., & Briggs, G. E. (1963). Effects of task complexity and task organization on the relative efficiency of part and whole training methods. *Journal of Experimental Psychology, 65*(3), 217.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. *Cognitive skills and their acquisition, 1*.

Noe, R. A. (1986). Trainees' Attributes and Attitudes: Neglected Influences on Training Effectiveness. [Article]. *Academy of Management Review, 11*(4), 736-749. doi: 10.5465/amr.1986.4283922

Norman, D. (1990). The'problem'with automation: inappropriate feedback and interaction, not'over-automation'. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 327*(1241), 585-593.

Norman, D. A. (1990). The'problem'with automation: inappropriate feedback and interaction, not'over-automation'. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 327*(1241), 585-593.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*, 429-429.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. [Article]. *Educational Psychologist, 38*(1), 63-71.

Paquette, G. (2001). Telelearning Systems Engineering - Towards a New ISD Model. [Article]. *Journal of Structural Learning & Intelligent Systems, 14*(4), 319.

Parasuraman, R. (1992). Theory and design of adaptive automation in aviation systems: DTIC Document.

Parasuraman, R., & Manzey, D. H. (2010). Complacency and Bias in Human Use of Automation: An Attentional Integration. [Article]. *Human Factors, 52*(3), 381-410. doi: 10.1177/0018720810376055

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced'complacency'. *The International Journal of Aviation Psychology, 3*(1), 1-23.

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors, 38*(4).

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making, 2*(2), 140-160.

Parasuraman, R., & Wickens, C. D. (2008). Humans: Still vital after all these years of automation. *Human Factors, 50*(3), 511-520.

Russell, S., Norvig, P., & Artificial Intelligence, A. (1995). A modern approach. *Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs*.

Schmidt, R. A., & Bjork, R. A. (1992). NEW CONCEPTUALIZATIONS OF PRACTICE: Common Principles in Three Paradigms Suggest New Concepts for Training. [Article]. *Psychological Science, 3*(4), 207-217.

Sheridan, T. (2000). Function allocation: algorithm, alchemy or apostasy? *International Journal of Human-Computer Studies, 52*(2), 203-216.

Sheridan, T. B. (1993). Space teleoperation through time delay: Review and prognosis. *Robotics and Automation, IEEE Transactions on, 9*(5), 592-606.

Sheridan, T. B. (1997). Trends in human control of aircraft, railway trains, and highway vehicles. In T. B. Sheridan & T. Van Lunteren (Eds.), *Perspectives on the human controller: Essays in honor of Henk G. Stassen.* (pp. 157-166). Mahwah, NJ Lawrence Erlbaum Associates Publishers.

Sheridan, T. B., & Parasuraman, R. (2000). Human versus automation in responding to failures: An expected-value analysis. *Human Factors, 42*(3), 403-407. doi: 10.1518/001872000779698123

Sheridan, T. B., & Parasuraman, R. (2005). Human-automation interaction. *Reviews of human factors and ergonomics, 1*(1), 89-129.

Sheridan, T. B., & Verplank, W. (1978). Human and computer control of undersea teleoperators (Man-Machine Systems Laboratory Report): Cambridge: MIT.

Singer, P. W. (Writer). (2009a). Wired for war: The robotics revolution and conflict in the 21st century, *The Penguin Press*.

Singer, P. W. (2009b). *Wired for war: The robotics revolution and conflict in the twenty-first century*: Penguin Press.

Smith, C. (2011, 04 November 2011). In Pakistan Drones Kill Our Innocent Allies, *New York Times*. Retrieved from http://www.nytimes.com/2011/11/04/opinion/in-pakistan-drones-kill-our-innocent-allies.html

Stammers, R. (1982). Part and whole practice in training for procedural tasks. *Human Learning, 1*(3), 185–209.

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. [Article]. *Educational Psychology Review, 10*(3), 251-296.

USAF, H. (2009). USAF Unmanned Aerial Systems Flight Plan 2009-2047: HQ USAF.

USAF/AFRL. (2002). *Predator STE (Synthetic Training Environment) Installation Disk v 2.0*. U.S. Air Force Research Laboratory

Van Merriënboer, J. J. G., Kester, L., & Paas, F. (2006). Teaching complex rather than simple tasks: balancing intrinsic and germane load to enhance transfer of learning. [Article]. *Applied Cognitive Psychology, 20*(3), 343-352. doi: 10.1002/acp.1250

Weisbin, C., Blitch, J., Lavery, D., Krotkov, E., Shoemaker, C., Matthies, L., et al. (1999). Miniature robots for space and military missions. *Robotics & Automation Magazine, IEEE, 6*(3), 9-18.

Wickens, C., & Colcombe, A. (2007). Dual-task performance consequences of imperfect alerting associated with a cockpit display of traffic information. *Human Factors, 49*(5), 839-850. doi: Doi 10.1518/001872007x230217

Wickens, C. D. (2009). ISAP Keynote THE PSYCHOLOGY OF AVIATION SURPRISE: AN 8 YEAR UPDATE REGARDING THE NOTICING OF BLACK SWANS.

Wickens, C. D., & Andre, A. D. (1990). Proximity compatability and information display: effects of color, space, and objectness on information integration. *Human Factors, 32*(1), 61-78.

Wickens, C. D., Dixon, S. R., & Ambinder, M. S. (2006). Workload and automation reliability in unmanned air vehicles. In N. J. Cooke, H. L. Pringle, H. K. Pedersen & O. Connor (Eds.), *Human factors of remotely operated vehicles.* (pp. 209-222). Amsterdam Netherlands: Elsevier.

Wiener, E. (1988). Cockpit automation. *Human factors in aviation, (A 89-34431 14-54). San Diego, CA, Academic Press, Inc., 1988*, pp. 433-461.

Wightman, D. (1983). Part-task training strategies in simulated carrier landing final-approach training: Training Techniques Development Branch, Human Factors Laboratory PE 63733N, Naval Training Equipment Center, Orlando FL 32813.

Wightman, D., & Lintern, G. (1985). Part-task training for tracking and manual control. *Human Factors, 27*(3), 267-283.