MITIGATING EFFECTS OF WORKING MEMORY CONSTRAINTS ON AUTOMATION USE THROUGH INTERFACE REDESIGN

by

Haneen Rezik Saqer
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Psychology

Committee:

_________________________________ Director

____

____

_________________________________ Department Chairperson

_________________________________ Program Director

_________________________________ Dean, College of Humanities and Social Sciences

Date: ___________________________ Summer Semester 2015
George Mason University
Fairfax, VA

Mitigating Effects of Working Memory Constraints on Automation Use through Interface Redesign

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Haneen Rezik Saqer
Master of Arts
George Mason University, 2011
Master of Health Administration
Washington University School of Medicine in St. Louis, 2003
Bachelor of Arts
Trinity University, 2001

Director: Tyler Shaw, Assistant Professor
Department of Psychology

Summer Semester 2015
George Mason University
Fairfax, VA

DEDICATION

This dissertation is dedicated to the memory of my beloved mother, Fadwa Ghazy Saqer, who selflessly supported my career change, cross-country move, and academic endeavor in the face of her ongoing health challenges.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Raja Parasuraman, for his mentorship, support, and encouragement throughout my doctoral studies. He left us much too soon and I wish he could have been here to share this accomplishment. I would also like to thank Tyler Shaw for graciously stepping-in to serve as chair of my dissertation committee during a difficult time for the psychology department and for himself personally. I am indebted to my committee members, Carryl Baldwin and Ewart de Visser, for their feedback and advice throughout the defense process. I am grateful to my family and friends who provided me with emotional support and cheerleading when I needed it most. My father, Rezik Saqer, encouraged me to pursue my dream even when it meant leaving my job as administrator of his medical practice. My husband, Seyamak Shahrzad, stood by me through the entire process and cheerfully made me coffee for many sleepless nights. He always believed in me more than I believed in myself. My best friend since our first day at orientation, Nicole Werner, served as an invaluable sounding board for both academic and personal decisions and patiently listened to many, many hours of venting. Finally, I owe many thanks to my sister Aseel Saqer, in-laws, aunts, uncles, and cousins who believed in me every step of the way. None of this was possible without each and every one of you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

MITIGATING EFFECTS OF WORKING MEMORY CONSTRAINTS ON AUTOMATION USE THROUGH INTERFACE REDESIGN

Haneen Rezik Saqer, Ph.D.

George Mason University, 2015

Dissertation Director: Dr. Tyler Shaw

This dissertation investigated the role of individual differences in human use of automation in a simulated command and control task. Using this knowledge we then sought to redesign the simulation interface to improve human-automation interaction. In the first study, participants completed a battery of cognitive tasks to measure working memory capacity, simple memory span, and controlled attention ability. They then performed a simulated air defense task under varying levels of workload and automation assistance. Eye tracking data recorded fixations to capture eye movements during completion of each scenario. Although individual difference measures correlated with primary task performance, they did not predict use of automation. Only average percent of fixations on the automation messaging interface correlated with automation use. Therefore, the second study introduced a redesigned automation interface with the integration of an auditory chime and a visual flicker to promote additional fixations to the message interface and encourage increased automation use. However, this redesign did not increase average fixation percentage and surprisingly resulted in lower use of automation. This finding emphasized Parasuraman and Riley's (1997) warning that automation can change user behavior in unintended ways. Another notable finding from the study is the unexpected result that short term memory predicted primary task performance. Further, this study provides evidence to support the use of eye tracking measures as a continuous unobtrusive measure of automation use in complex systems. Limitations and future research are also discussed.

INTRODUCTION

Automation

Substantial technological advances have significantly altered many facets of the human work experience in industries such as aviation, medicine, manufacturing and modern military operations. Specifically, the increased development and implementation of automation introduces a unique set of human factors implications. Automation is defined as a computer or machine that performs partially or fully a task previously performed by a human partially or fully (Parasuraman & Riley, 1997; Parasuraman, Sheridan, & Wickens, 2000). Tasks once solely completed by humans are increasingly aided or performed by automation devices. Many system designers and managers are tempted to employ automation whenever possible because of the perceived cost and safety features afforded by automation (Kaber, Omal, & Endsley, 1999; Lee & Seppelt, 2009; Parasuraman, Bahri, Deaton, Morrison, & Barnes, 1992; Parasuraman & Riley, 1997; Sarter, Woods, & Billings, 1997). Likewise, companies often hastily implement automation in products to make them more appealing and to be first in the market to promote such cutting-edge innovations. However, the limitations of automation are often overlooked and require careful consideration prior to implementation during the design process. Introduction of automation to a system is more than the addition of computers and technological devices; automation changes the nature of work performed by humans, often in unintended and unanticipated ways (Parasuraman & Riley, 1997). Thus, the benefits and limitations of automation should be carefully considered in the implementation of automated systems.

Benefits of Automation

When designed properly, automation decreases human workload, increases productivity, and improves safety, thereby improving overall efficiency. For example, Grabowski & Hendrick (1993) reported the number of crew members required to operate vessels like cargo ships and oil tankers decreased to 8-12 members from 30-40 members required 40 years ago. In the aviation industry, the use of automation has also greatly increased efficiency by reducing fuel consumption and flight times. During the late 1970s, the number of crew required to operate a commercial airliner was reduced from three to two, largely as a result of automation of flight engineering functions. Aviation automation has also improved safety by aiding crew members in flight operations during inclement weather conditions (Parasuraman et al., 1992). In certain instances, automation aids human operators with tasks they would not be able to perform in time-critical events. For example, while driving in wet or icy conditions automatic braking systems assist human operators by decreasing stopping distances (Parasuraman & Riley, 1997). Modern smart cars, equipped with adaptive cruise control, adjust driving speeds based on headway and time-to-collision (Lee & Seppelt, 2009). These benefits make automation an appealing option to include in system design.

Limitations of Automation

In addition to the aforementioned benefits, there are also potential drawbacks to automation that must be considered carefully prior to implementation during the system design process. These costs primarily occur because automation can be imperfect, for several reasons. Automated systems can also be poorly designed, and operators may also not receive appropriate training in using automation. The human performance costs of automation include, among others, increased mental workload, decreased situational awareness, skill degradation, and complacency (Parasuraman et al., 2000). Although one of the main goals of automation is to decrease workload, Wiener (1988) and Kirlik (1993) found that increased workload occurs in systems implementing "clumsy" automation. Examples of clumsy automation include systems in which the automation is difficult to initiate or engage and in instances where additional physical work, such as extensive data entry, is required to activate the automation. Another potential cost of automation is decreased situational awareness. When changes to the system are carried out by another agent, specifically automation, human operators become less aware of changes in the environment and system states. This is particularly true in systems implementing high levels of automation during decision-making functions in dynamic environments because human operators lose the ability to evaluate the raw data required to make a decision when automation fails (Parasuraman et al., 2000). Similarly, as humans become accustomed to using automation regularly, they rely upon manual and cognitive skills less frequently. As the length of nonuse increases, so too does the amount of skill decay (Arthur, Bennett, Stanush, & McNelly, 1998). This skill degradation becomes problematic in instances of automation failure which require human operators to revert to manual skills. Additionally, complacency arises in systems comprising high levels of automation due to overreliance. This misuse of automation can lead to unique human errors such as decision biases and failures in monitoring (Bahner, Hüper, & Manzey, 2008; Parasuraman & Manzey, 2010; Parasuraman & Riley, 1997). In systems possessing highly consistent but imperfect reliability, human operators exhibit excessive trust in automation, subsequently leading to monitoring failures. Parasuraman, Molloy, and Singh (1993) demonstrated complacency empirically in an automation failure detection task. Participants detected less automation failures in conditions with consistently reliable automation compared to a manual condition. Furthermore, this complacency effect was not seen in a condition with automation whose reliability was variable. This finding has severe practical implications given that most automated systems are highly, but not perfectly, reliable.

Design Considerations to Mitigate Limitations of Automation

To mitigate some of the potential pitfalls of automation, careful considerations regarding both the context of the task and the human operator should drive automation design decisions. Parasuraman, Sheridan, and Wickens (2000) proposed a model for determining the appropriate type and level of automation for any given function. The model offers a human-centered approach and evaluates the appropriateness of automation in the context of the human information processing model (as opposed to the more frequent technology-centered approach primarily concerned with cost and technical capability). By using the proposed model and iterative design process as guide, unforeseen costs of automation can be mitigated. Several studies have documented the differential benefits and costs of automation types and levels (for a recent meta-analysis, see Onnasch, Wickens, Li, & Manzey, 2013). Context-aware automation is an example of adaptive automation, which has been shown in many studies to reduce some of the costs of static automation (Feigh, Dorneich, & Hayes, 2012; Inagaki, 2003; Kaber & Endsley, 2004; Parasuraman, 2000; Scerbo, 2007; Sheridan, 2011). For example, Parasuraman, Mouloua, & Hilburn (1999) used varying degrees of tracking difficulty in a flight simulation task to manipulate level of operator workload. Pilots performed the task in three blocks with a workload profile of high-low-high, corresponding to the take-off, cruise, and landing phases of flight. The pilots were assigned to one of three kinds of automation: workload-matched, "clumsy-automation," and a control group. The researchers found that when automation assistance was matched to the required workload of a task, tracking performance significantly improved compared to a control group. Performance not only improved during matched trials, but also in subsequent tasks after the automation was removed. This study also empirically demonstrated the effect of "clumsy automation" (Wiener, 1988); when automation was mismatched to operator workload, performance benefits from automation were eliminated and in some cases performance decreased from baseline.

Kaber, Omal, & Endsley (1999) performed studies varying the levels of automation in a telerobot task to address the specific limitation of decreased situational awareness introduced by automation. While they did find enhanced performance and lower subjective workload in high levels of automation, intermediate levels of automation best supported situational awareness by keeping the human operator involved in the task. This study provided further evidence for the need to evaluate the level of automation not only in the context of optimal performance, but in consideration of minimizing the costs associated with automation. Finally, in regards to overreliance, Parasuraman, Mouloua, and Molloy (1996) demonstrated that implementation of adaptive task allocation can reduce complacency in human monitoring. In adaptive task allocation, the system transfers control of tasks between human and automated control based upon predetermined criteria. Participants performed a multi-task flight simulation under full manual control and were then randomly assigned to one of the following three automation groups: model-based adaptive, performance-based adaptive, or nonadaptive. In the model-based adaptive condition, control of a previously automated task was returned to all users in the fifth 10-minute block of the trial. Participants in the performance-based group only had automation removed if performance fell below 55% after 40 minutes of completing the task. The nonadaptive condition was fully automated throughout all trials, but participants were notified that the automation was unreliable and they would therefore need to monitor the automated system. Accuracy of detection failure improved when automation was varied throughout the completion of the task compared to the static automation condition. This adaptive task allocation not only mitigated inefficient performance due to complacency, but also significantly enhanced monitoring performance in postallocation phases compared to preallocation phases. Additionally, because the operators were required to perform the task manually at various points throughout the experiment, the chance of skill degradation significantly decreased, thereby also reducing the "out of the loop" performance costs associated with static automation designs.

Adaptive Automation (AA)

Adaptive automation (AA) can further mitigate the previously described automation limitations because it can be designed to be more flexible, context-dependent, and user-specific than functional task allocation (de Winter & Dodou, 2014). Responsibility of control can be altered in real-time so that the automation becomes more flexible to changes in operator state and the environment. In this way, automation can be invoked during the times it is most needed (Scerbo et al., 2001). For example, static automation may be associated with "out-of-the-loop" performance costs (Endsley & Kiris, 1995), including automation complacency and bias (Parasuraman & Manzey, 2010; Parasuraman et al., 1993). Conversely, if automation is only engaged during times of under-load, the operator may turn off the automation or become frustrated with the system. Invoking automation in an adaptive manner alleviates some of these concerns. For example, in a study examining human-robot supervision, Parasuraman, Cosenzo, and de Visser (2009) found that while mental workload decreased in both an adaptive automation condition and static automation condition compared to manual performance, workload decreased significantly more when automation was adaptive rather than static. Participants supervised both uninhabited air and ground vehicles (UAVs and UGVs) in addition to two related tasks in the context of a military reconnaissance mission. In the static automation condition, participants were consistently aided by an automated target recognition (ATR) system. However, in the adaptive automation condition, the ATR system was invoked only if participant performance fell below a predetermined accuracy threshold of 50%. The findings of this study support the notion that context-dependent automation (in this case operator performance dependent) not only enhances performance, but also reduces mental workload to a greater extent than static automation.

A study by Prinzel, Freeman, Scerbo, Mikulka, and Pope (2003) further supported the case for adaptive automation. They found similar results regarding improved performance and decreased mental workload, but also identified that participants in an AA condition had additional attentional resources available to allocate to a secondary task. Participants of the study completed a modified version of the NASA MAT (Multi-Attribute Task) battery in one of three conditions. The first condition employed adaptive automation (AA), in which the individual operator's EEG engagement index (20 beta/(alpha + theta)) was used to switch between manual and automated states. Automation was turned off (or remained in manual mode) as long as the engagement index was below the individual operators' baseline engagement index. Conversely, automation was turned on (or remained in automatic mode) if the engagement index was above baseline. Participants in a yoked condition were paired to participants in the adaptive automation condition. In the yoked condition, participants received the same schedule of automation mode switches as their automation counterparts to control for any effects of the specific patterns of task mode switches. In the last condition, participants experienced a random schedule of automation switches. ERP data was collected for all participants while performing a secondary auditory oddball task. Results indicated that participants in the AA condition demonstrated better tracking performance and lower NASA-TLX scores for subjective workload than participants in either the yoked or control conditions. Additionally, the P300 response for the ERP for AA participants was significantly larger than for participants in the yoked and control conditions. The working hypothesis driving the use of the ERP data was that the amplitude of the P300 elicited from the secondary auditory oddball task should be proportional to the attentional resources invested in the task. The significantly larger P300 amplitudes for the AA participants indicated that that they were able to free attentional resources to perform this task. These results are encouraging and suggest that adaptive automation not only optimizes mental workload and performance in primary task situations, but can also support the freeing of attentional resources to complete secondary tasks.

Wilson and Russell (2007) also provided empirical support for the advantages of adaptive automation and demonstrated the additional benefits of customizing automation to individual abilities. Participants were required to locate and designate targets using pre-established rules in a complex aviation task in which operators were responsible for four uninhabited air vehicles (UAVs). Several psychophysiological measures were synthesized into artificial neural networks (ANN) to determine the operator functional state (OFS), which would then modulate AA. There were four automation conditions manipulated: no adaptive aiding, adaptive aiding, random aiding, and leave-on aiding. In the adaptive aiding condition, if the ANN determined that the operator was in a state of high cognitive workload, the UAV task was modified to a lower difficulty task (either by decreasing the velocity of the vehicles or displaying vehicle health status messages and allowing for more time to complete target selection). Conversely, when the ANN detected low cognitive workload, the automation returned the task to the original level of difficulty (higher vehicle speeds). In the leave-on aiding condition, automation was turned on when the ANN determined that the operator had reached a high workload state, and remained on until the weapon release command was given. In the no aiding, adaptive aiding, and random aiding conditions, there was an additional experimental manipulation for difficulty of task. In half of the conditions, the difficulty of task (vehicle speed) was determined by setting the speed to either an individual mean speed or a group-derived mean speed. The results indicated that the performance improved in the adaptive aiding condition compared to no aiding and random aiding. Furthermore, improvement was greater when the task difficulty was adjusted to individual (as opposed to group-derived) criteria. This finding suggests that the customization of automation and difficulty level to the individual operator has even greater potential benefit than adaptive automation developed based on group performance means.

Adaptive Automation Invocation Strategies

As evidenced in the various experimental designs of the studies described thus far, there exist several strategies to invoke adaptive automation. Parasuraman et al. (1992) classify these strategies into five main categories: the presence of a critical event, operator performance, operator physiological assessment, operator modeling, and hybrid methods that combine one or more of these methods. Each of these invocation strategies possesses specific advantages and disadvantages which must also be evaluated prior to automation implementation. For example, in critical event invocation the implementation of automation is tied to the occurrence of specific tactical events such as a "pop-up" weapon delivery sequence that leads to the automation of all aircraft defensive measures (Barnes & Grossman, 1985). If the critical event does not occur, automation will not be employed. This kind of automation is relatively easy and inexpensive to implement; however, the inherent simplicity of such a method fails to account for actual operator workload or performance. Critical event invocation also requires that the events be anticipated which is rarely the case in dynamic complex environments.

In contrast, psychophysiological measures of operator workload offer much more sophisticated levels of flexibility and customization. As described in Scerbo et al. (2001), there are three main benefits to physiological assessments. First, physiological measures allow for continuous monitoring and, unlike behavioral measures, do not require overt responses. This allows for greater sensitivity in the measure, given that operators can exhibit similar performance metrics under varying task loads. Moreover, physiological measures are also diagnostic in that they provide additional inferential information when coupled with behavioral responses compared to behavioral responses alone. For example, the behavioral measure of reaction time may indicate working memory limitations or overload related to response-related processing. Coupled with reaction times, ERP data can help localize this overload to central processing. Lastly, physiological assessments not only provide information regarding when mental overload may be occurring, but also which brain networks may be affected.

Despite these superior advantages, psychophysiological assessments do possess their own set of limitations. Besides the generic limitations of cost and intrusiveness, psychophysiological measures have added considerations for both conceptual and technical sensitivity and diagnosticity (Scerbo et al., 2001). Another important consideration to implementing these measures is their intrusiveness. Many EEG and ERP studies conducted in laboratories require participants to remain stationary and silent. If these adaptive systems are to be implanted in real-world settings, these limitations are unrealistic and not easily complied with. Particularly with ERP, a secondary task is often necessary to evoke the desired potential. In applied situations, this secondary task adds to workload and may be shed by the operator, no longer contributing to logic of the adaptive model. Additionally, reliable psychophysiological measures should exhibit minimal amounts of artifact (noise) and be able to predict mental states across individual differences. Because one of the main benefits of adaptive automation is the ability to customize automation for individuals, psychophysiological measures must be compared to a control baseline calculated for each individual. Therefore, reliability can be difficult to obtain even within subjects because of outside influences like time of day, stress level, and medication usage (Byrne & Parasuraman, 1996).

Given these limitations of invoking AA based on real-time psychophysiological measures, new invocation strategies should be explored. One particular concern with traditional adaptive automation is that the frequent turning on and off of automation can lead to unpredictable workload. If adaptive automation shifts control of the system between human and machine based on criteria unknown to the operator, this can result in frustration in the operator and potential mistrust in the system. The inability of the operators to anticipate when they will be required to take back manual control and when automation will be invoked can become very unsettling for the user and can add to the cognitive load of the operator. In contrast, individual difference markers can be used to invoke automation. In this way, automation invocation remains customized to the individual, but results in less uncertainty for the human. Individual differences in working memory capacity (WMC) are of particular interest due their ability to predict performance in both basic perceptual tasks like the antisaccade and Stroop tasks (Kane, Bleckley, Conway, & Engle, 2001; Kane & Engle, 2003; Unsworth, Schrock, & Engle, 2004), as well as more complex cognitive tasks such as reading and multitasking (Bühner, König, Pick, & Krumm, 2006; Daneman & Carpenter, 1980). Working memory has also previously been implicated in the use of automation using command and control tasks (de Visser, Shaw, Mohamed-Ameen, & Parasuraman, 2010; McKendrick et al., 2013). Prior to discussing how individual differences in WMC can inform automation design, current models of working memory will be reviewed.

Working Memory Capacity (WMC)

Jarrold and Towse (2006) define working memory as the "ability to hold information in mind while manipulating and integrating other information in the service of some cognitive goal." This definition emphasizes the importance of both the storage and processing aspects of working memory. On the other hand, short-term memory only requires the *storage* of information for a limited amount of time. Although it is not uncommon to find the terms working memory and short-term memory used interchangeably, the distinctions between the two are important from both theoretical and practical perspectives. Incidentally, because the scoring of most working memory tests are spans (i.e. counts), many people refer to differences in working memory as differences in working memory *capacity*. However, it is important to realize that these span measures also reflect differences in processing speed. Currently, the most widelyaccepted theory of working memory is the multi-component model of Baddeley and Hitch (1974) and the revised model of Baddeley (2000). This model consists of two slave systems, the phonological loop and visuospatial sketchpad, which store and rehearse auditory/verbal and visual/spatial information, respectively. Both of these systems serve the central executive. Initially introduced as a catch-all component that encompassed general processing resources, the central executive is now believed to control such critical functions as focusing, dividing, and switching attention. In 2000 Baddeley added the last component to the model, the episodic buffer. This mechanism explained the integration of working memory information with long-term memory. Over four decades of research have shed light on the subcomponents of these systems, their neurological underpinnings, and their significance in special populations (i.e. children, Alzheimer's patients, and amnesiacs) (Baddeley, 2002, 2003, 2012; Repovs & Baddeley, 2006). In addition to the vast volume of research focused on the structure and function of the various components of working memory, other studies have explored the challenges of measurement and individual differences of working memory (for a review see Jarrold & Towse, 2006).

Measures of Individual Differences in Working Memory

Prior to the 1980s, several researchers attempted to measure working memory with simple short term memory tasks. However, these working memory scores weakly correlated with reading comprehension, despite the fact that many theorists argued working memory was essential to reading comprehension. In 1980, Daneman and Carpenter developed the first *complex* working memory task that taxed both the storage and processing components of working memory. Their reading span (RSPAN) task required participants to read sentences aloud and recall the final word of each sentence. Subjects were then tested on both reading comprehension and recall of the final words. This RSPAN measure accounted for differences in reading comprehension and correlated with verbal SAT scores; it is still used in present-day working memory studies. Daneman and Carpenter argued that differences in working memory were attributable to a domain general resource (i.e. shared between processing and storage aspects of working memory). In other words, they hypothesized that individuals with faster processing speeds had residual resources to serve the storage functions. This argument mirrors initial theories of attention and workload. Specifically, Broadbent's (1958) single-bottleneck model posited that cognitive tasks are completed in queue by a central processor (i.e. single resource). Just as the single resource theory of workload was later disputed, so too was this single resource theory for working memory.

In another model of individual differences of working memory, Engle and colleagues argued that not only is working memory unitary, but it is domain-free (Engle, Kane, & Tuholski, 1999; Engle, 2002; Turner & Engle, 1989). Essentially, they defined working memory as short-term memory with the addition of controlled attention: WM = STM + controlled attention. They postulated that controlled attention is the primary reason that WM span tasks differ from simple span tasks, and why WM tasks are correlated to higher cognition such as intelligence tests, SAT scores, and academic achievement. In fact, they argued that WM is only necessary in tasks that require controlled attention such as in situations where irrelevant information competes for attention or in situations of competition resolution (Engle, 2002). Their model comes from years of correlational studies investigating variation in working memory capacity (WMC). Following the work of Daneman and Carpenter's (1980) RSPAN, Turner and Engle (1989) investigated if the processing component of the complex span task (verification of sentences in RSPAN) needed to involve reading in order to correlate with reading comprehension (i.e. is WMC task dependent?). They concluded that it is not; the computation of mathematical operations taxed processing enough to provide a measure of WMC that correlated with RSPAN and reading comprehension. They called this measure OSPAN. OSPAN has been shown to correlate with a multitude of tasks ranging from antisaccade (Kane et al., 2001; Unsworth et al., 2004), Stroop (Kane & Engle, 2003) and allocation of visual attention (Bleckley, Durso, Crutchfield, Engle, & Khanna, 2003). They argued that because these tasks are dominated by basic attention (i.e. do not involve storage or verbal processing), WMC is the same thing as controlled attention (Kane et al., 2001). However, this theory may be oversimplifying the role of the phonological loop. Although Turner and Engle (1989) found that mathematical operations taxed processing enough to provide a robust measure of WMC, others argued that OSPAN was not a pure measure of controlled attention and made the case for domain-specific resources.

Shah and Miyake (1996) provided evidence for domain-specific resources for verbal and visuo-spatial information and a dissociation between storage and processing functions. They argued that although Turner & Engle (1989) were able to use arithmetic computations in the development of OSPAN, this did not remove a verbal storage component from the measure. Successful completion of the mental arithmetic tasks requires verbal coding. Therefore, OSPAN was not a pure measure of controlled attention. To provide evidence for the importance of domain-specific resources, they used the RSPAN (Daneman & Carpenter, 1980) to assess language processes because it taxed both storage and processing strictly with verbal information. Following this paradigm, they developed a spatial task that taxed both the storage and processing of spatial information using only spatial stimuli. Subjects were required to perform a mental rotation while simultaneously performing a spatial orientation tracking task. As expected, they found that RSPAN measures correlated highly with verbal ability and that spatial working memory scores correlated highly with spatial ability. More importantly, they found that the RSPAN measures did *not* correlate with spatial ability and spatial working memory scores did *not* correlate with verbal ability. Clearly this finding suggested the existence of separate resource pools for verbal and spatial information. They then conducted a follow-up experiment to differentiate between the storage and processing aspects of working memory. In this study they used an interference paradigm to cross processing and storage demands of both span tasks. They found that although the storage requirement was critical to the successful completion of a span task, the processing capabilities of an individual explained variance in performance above storage capacity alone. Taken together these findings suggest that not only do "two separate pools of domain-specific resources" exist, but these resources support both storage and processing aspects of task-specific performance.

This finding was further supported by research conducted by Bayliss, Jarrold, Gunn, and Baddeley (2003). Similar to the Shah and Miyake interference study (1996), they developed complex span tasks that crossed verbal and spatial processing with verbal and visuospatial storage. They found that both processing efficiency and storage capacity *independently* affected performance on these complex working memory tasks. More specifically, their results were consistent with a model consisting of one domain-general processing resource pool and two domain-specific storage pools (i.e. at least three separate pools). (These results were consistent in populations of both young and old participants.) Bayliss et al. (2003) attributed any residual variance in performance not explained by one of these three resources to the ability to coordinate storage and processing, supporting the existence of a central executive. This further supported work by other researchers whose statistical analyses produced three factors to explain performance on working memory tasks involving different modalities (Ackerman, Beier, & Boyle, 2002; Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002). Aside from providing converging evidence for the existence of three separate resource pools for verbal storage, visuospatial storage, and domain-general processing (i.e. a central executive), these studies also highlighted the substantial amount of individual differences in each of these components and their predictive power. Therefore, when examining the role of individual differences in WMC in regards to automation, it is critical to explore individual differences in each of the three subcomponents of working memory.

Finally, the controlled attention model of working memory does not account for effects attributed to the phonological loop. Namely, studies manipulating acoustic similarity of letters and words have shown that when stimuli are similar in sound, the rehearsal process suffers and the stimuli are more difficult to recall serially than when they share other similarities like meaning. This points to the acoustic component of rehearsal (Baddeley, 2002, 2003, 2012). The word-length effect also provides support for the phonological loop. Words of longer length are more difficult to recall than shorter words because as the time to rehearse each word approaches the two-second limit, words are forgotten before they can be rehearsed again (Baddeley, 2000, 2002). Further evidence can be found in the effect of articulatory suppression; when given a secondary task that prevents rehearsal (such as requiring participants to recite the word "the" continuously), retrieval declines greatly. Articulatory suppression also negates the word-length effect; when rehearsal is suppressed even short words are difficult to remember (Baddeley, 2000, 2002). Capacity of the phonological loop is limited and, although research findings differ regarding the exact number, most accepted estimates are in the range of between four and seven chunks of information (Miller, 1956).

Working Memory and Attention

Despite its inability to fully explain all components of working memory, the controlled attention model proves useful in exploring individual differences in executive function via visual attention tasks. For example, Bleckley et al. (2003) used a visual allocation task to show that low and high working memory individuals allocated attention differently. Low working memory individuals allocated attention as a spotlight, whereas high working memory counterparts allocated attention more flexibly. A similar task, the useful field of view task (UFOV®; Visual Awareness, Inc.), was developed to measure functional vision, divided visual attention, and selective visual attention. Functional vision is defined as the visual area from which information can be acquired without moving the eye or turning the head (Sanders, 1970). The UFOV® has predominantly been used to assess useful field of view in older adults (Ball, Beard, Roenker, Miller, & Griggs, 1988) and assess driving risk (Ball, Owsley, Sloane, Roenker, & Bruni, 1993; Ball & Owsley, 1993; Owsley, Mcgwin Jr, & Ball, 1998; Owsley, Ball, et al., 1998). However, it is also useful in measuring divided and selective attention. According to the controlled attention theory, individuals with higher working memory should exhibit superior performance in the divided and selective attention portions of the UFOV® task. Thus, we are including the UFOV® measure in our study to determine if differences in working memory span correlate to differences in visual attention allocation and if these differences can predict visual allocation in applied settings (i.e. automation interface).

In addition to control of visual attention, Kane and colleagues argue that the critical component of working memory is executive control, particularly in the presence of conflict (Engle & Kane, 2004; Engle, 2002; Kane & Engle, 2003). They propose a two factor model of controlled attention. Evidence for this theory comes from performance data on Stroop tasks. In order to successfully complete a Stroop task, participants must maintain two subgoals: 1) maintain the goal to respond to font color and 2) resolve the conflict presented by the word spelling. Kane and Engle (2003) found that low span participants exhibited faster reaction times in congruent trials compared to neutral trials than their high working memory counterparts. The authors considered this facilitation effect to reflect goal neglect; lower span participants found it more difficult to ignore word meaning. Additionally, low working memory individuals exhibited greater Stroop interference than their high span counterparts as manifested by lower accuracy in incongruent trials compared to neutral trial baselines. The authors contend that this interference effect represents the participants' attentional inability to resolve conflict. Therefore, we are incorporating this task in our study to determine if Stroop task performance correlates with working memory span and the attention abilities of participants using the automation interface.

Individual Differences in Automation Use

Many individual differences have been shown to affect human interaction with automation. Because these traits can be easily assessed in a cost-effective and unobtrusive manner they can inform automation design. Examples of traits that can contribute to performance with automation include: propensity for complacency (Singh, Molloy, & Parasuraman, 1993a, 1993b), trust in automation (Merritt & Ilgen, 2008), extraversion (Syrdal, Lee Koay, Walters, & Dautenhahn, 2007), neuroticism (Szalma & Taylor, 2011), working memory capacity (de Visser et al., 2010; McKendrick et al., 2013; Saqer & Parasuraman, 2014), control of executive function (Chen & Barnes, 2012; Chen & Terrence, 2009; Parasuraman & Manzey, 2010), including genetic markers (Parasuraman et al., 2013), spatial ability (Chen & Barnes, 2012; Chen & Terrence, 2009; Lathan & Tracey, 2002) and video game experience (Chen & Terrence, 2009). For example, in two studies of human-unmanned vehicle (UV) interaction, Chen and Barnes (2012) showed the importance of operator spatial ability in supervisory control proficiency. In tasks that required visual scanning, participants with greater spatial ability consistently outperformed their low-spatial-ability counterparts.

Specifically, working memory and genetic markers of executive function have been shown to predict supervisory control performance both in individual and team settings (Ahmed et al., 2014; de Visser et al., 2010; McKendrick et al., 2013; Parasuraman et al., 2013; Saqer & Parasuraman, 2014). These studies showed that operators and teams with higher working memory capacities and COMT Met/Met genotypes exhibited better performance than their counterparts. Furthermore, McKendrick et al. (2013) found that team working memory scores interacted with task load, suggesting that working memory impacts performance in high task load conditions to a greater extent than in low task load conditions.

Preliminary Findings

In another study using a supervisory control UV task, Saqer and Parasuraman (2014) found that although working memory scores did not correlate with overall performance, working memory was predictive of automation use in conditions in which the level of automation was mismatched to level of workload. In this task participants interacted with automation comprising of two components representing different stages and levels of automation. Participants performed an air defense UV simulation under conditions of high and low task load and were given zero, one, or two autonomous aids to assist them. These aids functioned as robotic teammates that patrolled the airspace and engaged enemies autonomously. The autonomous aids in this study represented action implementation stage of automation functioning at the highest level of automation. The actions of the aids were communicated to the participants via a text messaging system approximately five seconds prior to their occurrence (information analysis stage of automation presented at a low level). However, nothing prevented operators from engaging these targets on their own. Because the actions of the autonomous aids were preprogrammed, we were able to calculate an automation effectiveness measure reflecting the extent to which an operator coordinated her activity with the autonomous aids. This was calculated as: number of successful engagements completed by the autonomous aid(s) divided by the total number of pre-programmed engagements. Deviation from 100% effectiveness occurred when operators engaged the enemies unnecessarily (i.e. automation interference).

Findings indicated that in context-matched conditions (low workload with low level of automation and high workload with high level of automation) automation effectiveness did not correlate with working memory score. However, in the non-context-matched conditions, working memory score was positively correlated with use of automation. In other words, when automation was not adaptive, operators with low spans used the autonomous aids to a lesser extent. This means that they were less able to adjust during periods of automation that were too high or too low for the situation. It should be noted that overall performance between low- and high- working memory operators did not differ. To keep overall performance high while making less use of the autonomous aids, low span operators must have engaged more targets manually, effectively duplicating efforts and performing both their own work in addition to the work of the aid(s). Higher working memory operators most likely benefited due to increased ability to process the verbal texts communicated though the automation messaging system. It remains unclear if the disuse of the aids by low span operators was due to a suboptimal attention allocation strategy causing them to miss verbal texts from the messaging interface or if they perceived the messages but were unable to incorporate the information into their attack strategies.

Study Rationale and Hypotheses

Given these findings, the current series of experiments sought to further explore individual differences in working memory and their effects on automation use. As discussed previously, individual differences in working memory can result from differences in executive control (i.e. via the central executive) or differences in capacity of the slave systems, the phonological loop and visuospatial sketchpad. Due to the verbal nature of the automation messaging interface used in this experimental paradigm, we were specifically concerned with differences in the phonological loop. Once the specific causes of automation disuse among low span individuals are identified, interventions for these causes can be developed and tested. In addition to informing automation design, these findings can also help with personnel selection and training. For example, managers may want to consider pairing individuals with lower working memory scores with teammates or provide additional training using the automation interface.

This dissertation focuses on two research questions:

1. Is the disuse of automation exhibited by low-span participants due to attention allocation strategies governed by the central executive or due to overload of the phonological loop prohibiting the integration of new verbal information into current strategy?

2. Can automation be specifically redesigned to address each of these hypothesized deficiencies and mitigate automation disuse?

We hypothesized that differences in attention allocation (i.e. the central executive) would be the primary reason for automation disuse. Parasuraman and Manzey (2010) suggest that complacency is a manifestation of attentional processes. They argue that automation complacency occurs under conditions of multiple-task load when the operator fails to monitor the system at a sufficient level. Similarly, we suggest that under conditions of task overload, users shed the task of monitoring the automation, in this case the messaging interface. We believe this is particularly true when the automation is designed as an aid and there are no direct negative consequences of automation disuse. Furthermore, evidence for a genetic marker of attention has already been shown to influence performance in supervisory command and control tasks. Individuals with lower working memory exhibited both lower primary task performance as well as reduced verification rates for incoming information provided by the automation (Parasuraman et al., 2013). We also hypothesized that automation redesign can mitigate these individual differences in working memory. With proper redesign we believed participants could improve in level of automation use comparable to their high-span counterparts.

Planned Analyses and Results

The same performance measures used in the preliminary study were calculated for studies 1 and 2: red zone protection, attack efficiency, and automation effectiveness. Repeated measures analysis of variance (ANOVAs) were conducted for both studies to determine effects of task load and level of automation. As with the preliminary study, we did not expect to observe differences in red zone protection or attack efficiency based on working memory scores.

For the first study, individual difference measures for working memory (Aospan), eye tracking data, divided and selective attention (Useful Field of View; UFOV®), controlled attention (Stroop task performance), and phonological loop capacity (simple digit span) measures were regressed onto automation effectiveness to determine the predictive power of each. As with the previous studies, we expected general WMC measures using Aospan to correlate positively with automation effectiveness, particularly in non-context matched conditions. According to our hypothesis, we expected eye tracking data and attention allocation measures (UFOV®, Stroop task performance) to correlate with automation effectiveness. From the eye tracking data we hypothesized that individuals exhibiting more fixations on the messaging interface would use the autonomous aids to a greater extent. We also hypothesized that participants with longer mean gaze durations on the messaging interface would exhibit greater disuse of the autonomous aids. (According to Fitts, Jones, and Milton (1950), longer gaze durations reflect difficulty extracting information from the display.) Likewise, participants who exhibited superior divided and selective attention (UFOV® task) and controlled attention (Stroop task) would exhibit greater automation effectiveness. We did not expect a simple measure of differences in phonological loop capacity (simple digit span task) to correlate with automation effectiveness. While it is theoretically possible that differences in phonological loop can explain the individual difference effects observed in the preliminary study, this measure offers much less variability in healthy young adults. It has frequently been reported that short term memory capacity is in the range of seven plus or minus two items (Miller, 1956). The limited sensitivity of this measure makes it difficult to find an effect with the sample sizes of these studies. Nevertheless, we measured the short term memory span of participants to test this hypothesis. Furthermore, we implemented a partial-credit unit scoring scheme on this measure to allow for more granularity in performance (for methodological review, see Conway et al., 2005).

In the second study, subjects performed the simulated air defense tasks with a redesigned automation interface. The nature of the redesign was determined from the results of the first study. For example, if divided and selective attention abilities correlated with automation use, a redesigned interface could incorporate visual and auditory cues to assist lower span participants with dividing attention. Similarly, if differences in automation use were related to differences in simple verbal spans, a redesign could alter the nature of the message to spatial graphics. We hypothesized that the attention allocation redesign of the interface would stabilize automation effectiveness measures across all levels of WMC. In other words, the redesign should bring the level of automation use for low working memory participants to the level of their high working memory counterparts. Although a verbal redesign could theoretically improve automation use from baseline, it could also have a detrimental effect by increasing workload. Given that supervisory command and control tasks are already visuospatial in nature, this redesign could place additional burden on the visuospatial sketchpad leading to decreased primary task performance.

STUDY ONE

Method

Participants

Thirty-seven George Mason University undergraduate students (19 men and 18 women) with an average age of 21.1 (*SD* = 3.77) years participated in the study for course credit.

Apparatus

Participants used a desktop computer running Microsoft Windows XP connected to a 32-inch screen to complete a supervisory control task administered via the Distributed Dynamic Decision-making 4.0 simulation software (DDD®; Aptima, Inc.). The software simulated an air defense task in which operators protected a designated zone while simultaneously engaging incoming targets, (see Figure 1). Eye tracking data was collected via Tobii X60 eye tracking system that sampled at a rate of 60 Hz. The messaging interface was marked as an area of interest (AOI) for eye tracking metrics (see Figure 2).

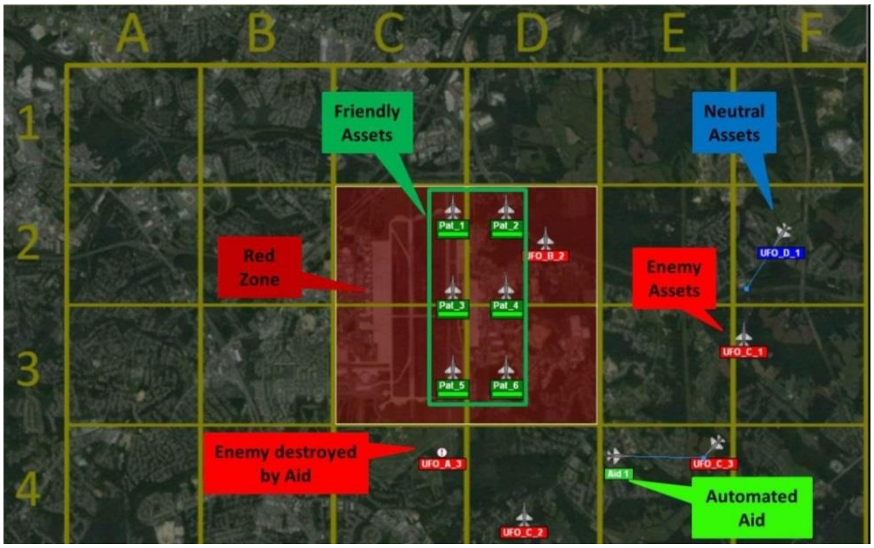

Figure 1. DDD® simulated air defense task with six unmanned vehicles controlled by operator (dark green) and one automated aid acting autonomously (light green). Operators used these assets to protect a no-fly red zone (shaded in red). Incoming enemy aircraft are displayed in red and neutral assets are indicated in blue.

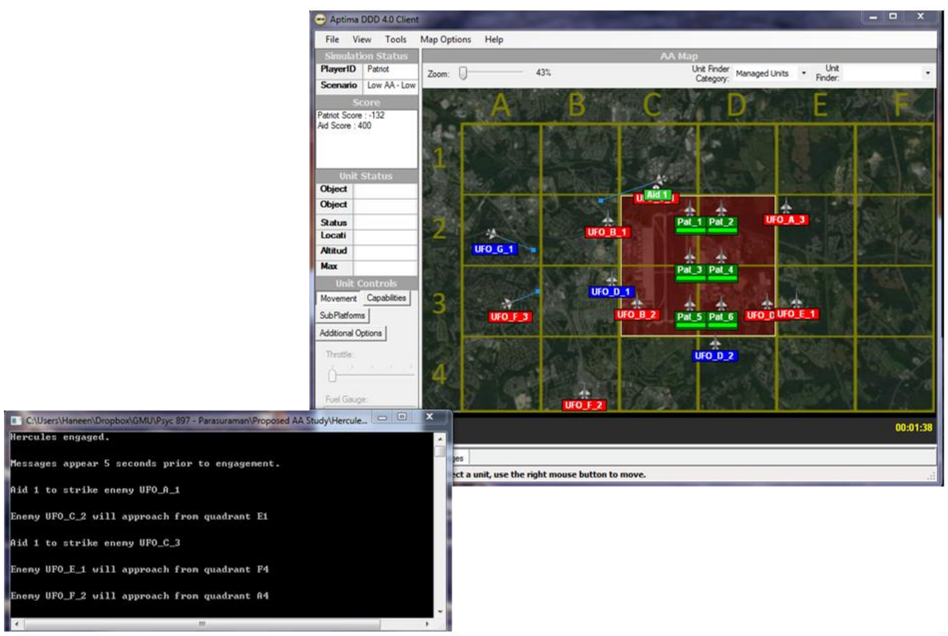

Figure 2. DDD® simulation with messaging interface in lower left corner. Messages appeared five seconds prior to an event's occurrence in the simulation.

DDD® Simulation

Operators were given control of six friendly unmanned assets to defend a no-fly "red zone" from incoming enemy and neutral targets. The enemy and neutral targets approached the red zone from multiple directions. Enemy targets always appeared with red labels whereas neutral targets were labelled in blue. Neutral targets had a 50% probability of changing to enemy targets which occurred without cue. The color of the aircraft label changed from blue to red to signify this change.

Participants were instructed to maximize their score for each scenario by achieving each of the three priorities: 1) engage and destroy all enemy targets; 2) prevent the incursion of any enemy targets into the red zone; and 3) avoid friendly fire against own friendly assets, autonomous aid assets, and neutral assets. To control the friendly assets, participants right-clicked the asset via mouse and selected a way-point on the map via left-click to indicate the desired travel path. To engage enemy aircraft, participants moved friendly assets near the desired target, activated missile capability via a menu on the left side of the screen, and right-clicked the target. If the enemy aircraft was within the range of the missile, the target was destroyed within 3s of engagement and disappeared from the display.

Automation in these scenarios comprised of two parts: an information messaging interface and the autonomous aids. The messaging interface was a script that ran concurrently with the simulation and delivered 25 relevant messages to the operators throughout the scenario. The messages contained the name and initial location of incoming enemy assets or the status of an autonomous aid. The messages served to help direct attention of the operator to immediate and imminent threats. Operators did not need to read every message to successfully complete the task; however, if an operator missed a message regarding the intention of the autonomous aid, it could have resulted in the operator engaging the same target as the aid, thereby duplicating efforts. The messages were delivered one line at a time 5s prior to the occurrence of the event. On average, messages contained approximately seven words (see Figure 2). This messaging interface was adapted from previous studies using the same paradigm (Ahmed et al., 2014; de Visser et al., 2010; McKendrick et al., 2013; Parasuraman et al., 2013). However, unlike in previous experiments that varied reliability of the messages, this study presented messages that were 100% reliable.

Aospan Task

Aospan was administered to derive a measure of overall working memory capacity. Participants performed the automated version of the OSPAN task, Automated OSPAN (Aospan; Unsworth, Heitz, Schrock, & Engle, 2005). This task presented participants with a simple arithmetic problem which they indicated as true or false immediately followed by a letter (e.g. (1*2) + 1 = 3; True or False; P). Participants recalled the letters in serial order at the completion of each set of trials. Set sizes ranged from three to seven letters and were presented in random order in 3 blocks resulting in a total of 75 total letters for recall. The Aospan program automatically recorded the working memory score for each participant at the conclusion of the task.

Useful Field of View (UFOV®)

Participants completed the useful field of view task (UFOV®; Visual Awareness, Inc.) to measure functional vision, divided visual attention, and selective visual attention.

The test was administered via computer and mouse and comprised of three subtasks presented in order of increasing difficulty. In the first task subjects identified a target presented in the center of the screen for varying lengths of time. In the second task, the participant detected the centrally located target while also localizing another object in the periphery. The final task was similar to the second but included the addition of distractors in the periphery. The UFOV® software automatically recorded scores for each of the three subtasks.

Stroop Task

To obtain a measure of controlled attention we conducted a color Stroop task. The task was written and administered via the Psychology Experiment Building Language (PEBL, (Mueller, 2012)). The task presented congruent, incongruent, and neutral word pairings. Congruent word pairs displayed the word in the matching hue (RED in red). Incongruent trials were combinations of the words RED, BLUE, or GREEN displayed in a mismatching hue (e.g. BLUE in red). Incongruent trials were combinations of the strings JKM, XTQZ, or FPSTW in the hues red, blue, or green. Trials were modeled after Kane and Engle's (2003) Stroop paradigm in which the task consisted of 75% congruent trials. A total of 288 trials were presented in three blocks. Each block contained 36 critical trials with equal numbers of neutral, incongruent, and congruent color-word pairings. Each block also contained 60 non-critical trials which were all congruent word pairs. Participants were instructed to complete the task as quickly but accurately as possible. Reaction times and accuracy measures were recorded automatically for each participant.

Simple Digit Span

To obtain a measure of short-term verbal memory (i.e. storage component of phonological loop), we presented sets of digits ranging from three to ten items on a computer screen. Trials were presented via PEBL software (Mueller, 2012). Digits were presented aurally via audio file at the same time that they appeared on the screen for 1 s with an interstimulus interval of 150 ms. At the conclusion of each set, participants were asked to recall the digits for that set in serial order. Sets of each size were presented three times in order of increasing size, resulting in a total of 24 sets. Participants received one point for any digit they correctly recalled. Maximum digit span score was 156 points.

Procedure

After reviewing and signing an informed consent, participants completed the Aospan, UFOV, Stroop, and simple digit span tasks. To gain familiarity with the DDD® task and understand the performance objectives, participants viewed a slide presentation. Participants then completed two practice trials: low task load with low level of automation (one autonomous UV) and high task load with high level of automation (two autonomous UVs), presented in counterbalanced order. During these trials participants practiced moving assets, engaging enemies, reading messages from messaging interface, and interacting with the autonomous aids. Finally, participants performed each of the four experimental trials presented in random order.

Experimental Design

A 2 x 2 repeated-measures factorial design was used to vary task load (low, high) and level of automation (low AA, high AA). Low task scenarios contained 60 enemy and neutral targets whereas high task load scenarios contained 75 total aircraft (see Table 1).

Workload was determined by the degree of visual, cognitive, and motor task load required by the operators to successfully complete task goals. Users were unaware of which enemies would switch to neutrals and which would reach the red zone. This required them to carefully monitor all UVs as they simultaneously coordinated their strategies with the autonomous aids. Therefore, the total number of UVs in the scenario was used to determine low and high task load, not the number of red zone incursions. Approximately 80% of total aircraft were enemy targets (including neutral UVs that converted to enemies). Because each scenario was seven minutes in duration, enemy targets appeared at a rate of one every nine seconds in the low task load conditions and one every seven seconds in the high task load conditions. Targets entered the scenario from the map perimeter and moved in randomly programed routes toward the red zone located in the center of the playfield. Targets completed two flight paths prior to disappearing. Automation levels varied by the number of autonomous aids available in the scenario: 1 or 2. Aids were friendly UVs that autonomously travelled around the red zone and engaged pre-programmed targets in their immediate proximity. Operators were informed of these engagements 5s prior to their occurrence via the messaging script; however, nothing prevented operators from engaging these targets by themselves. Each aid was programmed to eliminate 20% of enemy targets. In scenarios with one aid, the UV travelled in a circular counterclockwise pattern. In scenarios with two aids, each UV was responsible for half of the air space, one for the left and one for the right.

Table 1. Experimental Design of UVs Programmed in Each Scenario

Level of Automation	Total UVs	Neutral UVs	Neutral UVs Converted to Enemies	Enemies	Enemies Engaged by Aids	Enemies Programmed to Penetrate Red Zone with 100% Aid Use	Enemies Programmed to Penetrate Red Zone with 0% Aid Use
Low Task Load							
Low – 1 Aid	60	11	9	40	10	14	17
High – 2 Aids	60	11	9	40	20	10	14
High Task Load							
Low – 1 Aid	75	13	12	50	12	9	10
High – 2 Aids	75	13	12	50	25	8	13

Dependent Measures

The play logs for each scenario recorded all aircraft movements, engagement attempts, successful engagements, and incursions to the red zone. Engagement attempts were recorded each time the operator activated a missile. Unsuccessful attacks resulted from participants selecting targets out of range of the missile or execution errors in activating and launching missiles (i.e. inappropriate mouse-clicks and menu selections within the simulation interface). Using this data, we calculated the following performance measures: 1) red zone protection: 1 – (number of enemy aircraft that penetrated the red zone/total number of enemy aircraft programmed to penetrate the red zone); 2) attack efficiency: successful enemy engagements/total engagement attempts; 3) automation effectiveness: successful aid engagements/total programmed engagements. These measures were derived from previously published studies using the same simulation software (Ahmed et al., 2014; de Visser et al., 2010; McKendrick et al., 2013; Parasuraman et al., 2013). Each scenario differed in the total number of enemy targets and the total number of enemy UVs that were programmed to penetrate the red zone (see

Table 1). Therefore, all measures were computed as percentages ranging from 0 to 1. For attack efficiency and red zone safety, higher values indicate superior performance. For the automation effectiveness measure, lower values reflect disuse of the automation (Parasuraman & Riley, 1997). Unsuccessful engagements of the aids only occurred when operators engaged the enemies unnecessarily. Standard errors of the mean were calculated as standard deviations divided by square root of the sample size for each measure.

Eye tracking measures. Areas of interest (AOIs) were established around the messaging interface and the simulated playfield; they were defined as X and Y coordinates on the display. Two eye tracking measures were calculated for each participant for each experimental scenario: fixation count and mean gaze duration. Fixation count was calculated as the number of fixations within each AOI. More fixations on the messaging interface AOI indicated greater importance (Fitts et al., 1950). Therefore, higher fixation counts reflected more attention allocation to the automation and are hypothesized to correlate with automation use. Mean gaze duration was calculated as the cumulative length of consecutive fixations in the messaging AOI including the relatively small amount of time for short saccades between these fixations. Fitts et al. (1950) predicted that longer gaze durations reflected difficulty extracting information from the display. Thus, we hypothesized that longer mean gaze durations in the messaging AOI would be correlated with automation disuse. We were also interested to see if longer gaze durations correlated to lower simple span. If participants experienced difficulty with verbal information, it would be helpful to determine if low simple verbal span is a contributing factor.

Results

Primary Performance Measures

Red zone protection. To assess the effect of task load and level of automation on red zone protection, we conducted a 2 x 2 repeated measures ANOVA. The analysis revealed a marginal effect of task load $F(1, 36) = 3.16$, $p = .08$, $\eta_p^2 = .081$. Participants more successfully defended the red zone in low task conditions ($M = 47.0\%$, $SEM = 2.40\%$) than in high task load conditions ($M = 42.1\%$, $SEM = 1.90\%$). There was also a main effect for level of automation, $F(1, 36) = 18.21$, $p < .001$, $\eta_p^2 = .336$. Participants exhibited better protection of the red zone with one automated aid ($M = 49.0\%$, $SEM = 1.90\%$) than they did with two aids ($M = 40.0\%$, $SEM = 2.00\%$). Higher level of automation differentially affected red zone protection in the low and high task condition, as evidenced by a significant interaction, $F(1, 36) = 22.80$, $p < .001$, $\eta_p^2 = .388$. In the high task load condition, one aid was helpful ($M = 51.4\%$, $SEM = 2.81\%$) whereas two aids resulted in reduced red zone protection ($MD = 32.8\%$; $SEM = 2.29\%$) (see Figure 3).

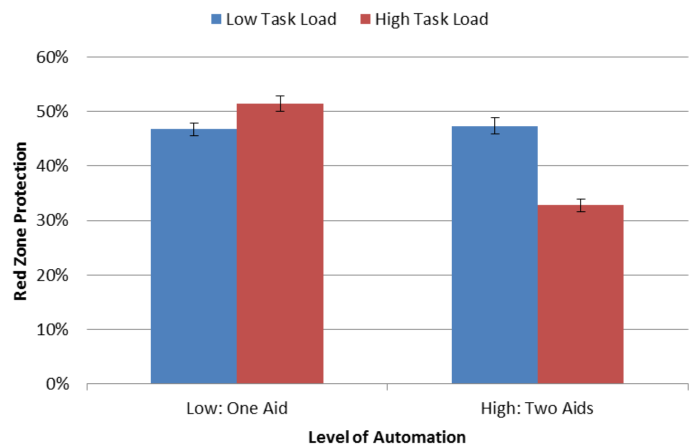

Figure 3. Study One: Mean red zone protection percentage for low and high task load conditions as a function of level of automation (low, high). Error bars represent standard errors of the mean.

Attack efficiency. Attack efficiency, a primary performance measure reflecting a participant's ability to successfully engage enemies in relation to engagement attempts, was also subjected to a 2 x 2 repeated measures ANOVA. Participants exhibited more efficient engagements in low task conditions (M = 74.6%; SEM = 1.70%) compared to high task conditions (M = 71.0%; SEM = 1.50%), as evidenced by a significant main effect of task load $F(1, 36) = 10.20$, $p < .01$, $\eta_p^2 = .221$. Level of automation did not influence attack efficiency, $F(1, 36) = 0.01$, $p = .92$. The analysis did not reveal an interaction between level of task load and level of automation, $F(1, 36) = 2.80$, $p = .10$ (see Figure 4).

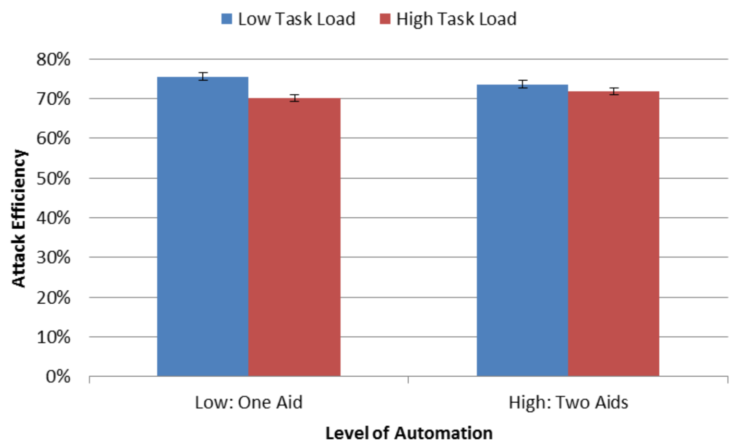

Figure 4. Study One: Mean attack efficiency for low and high task load conditions as a function of level of automation. Error bars represent standard errors of the mean.

Automation effectiveness. To determine how level of task load and level of automation influenced participants' use of the automated aids, we conducted a final 2 x 2 repeated measures ANOVA on the automation effectiveness measure. Level of task load did affect use of the autonomous aids, as reflected by a significant main effect $F(1, 36) = 4.31$, $p = .05$, $\eta_p^2 = .107$. Participants used the automated aids more in low task conditions ($M = 68.2\%$; $SEM = 2.50\%$) than they did in high task load conditions ($M = 65.0\%$; $SEM = 2.90\%$). Level of automation did not influence a participant's use of automation, $F(1, 36) = 0.06$, $p = .81$. However, there was a significant interaction between level of task load and level of automation, $F(1, 36) = 23.16$, $p < .001$, $\eta_p^2 = .391$. In the high task conditions participants used the automated aids more when they were only given one aid ($M = 69.8\%$; $SEM = 3.27\%$) than they did when they were given two aids ($M = 60.2\%$; $SEM = 3.05\%$) (see Figure 5).

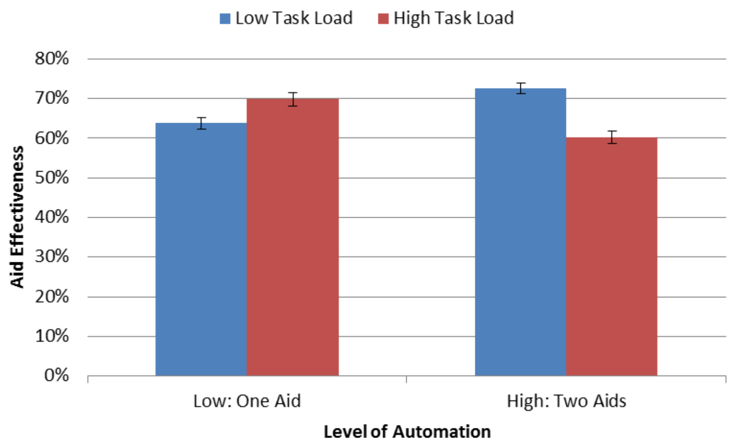

Figure 5. Study One: Mean use of automation for low and high task load conditions as a function of level of automation (low, high). Error bars represent standard errors of the mean.

Individual Difference Measures

Aospan. Automated operation span reflected the number of correctly recalled items in all presented sets, with a maximum possible score of 75. Aospans for participants ranged from 0 to 70 (M = 36.70, SD = 20.59). Surprisingly, participant Aospan scores only correlated with simple digit spans, $r(35) = .52$, $p < .01$ (see Table 2).

Useful Field of View (UFOV®). Performance on the second UFOV subtask represented divided attention ability. Scores below 100 are considered within normal range. All but one participant exhibited normal divided attention ability (M = 31.1, SD = 60.3*)*. For selective attention ability, scores below 350 on the third subtask are considered normal. All but one participant exhibited normal selective attention (*M = 50.4, SD =* 81.6*)*. Deficits in visual attention on the UFOV task are most common in older adults. Because we sampled a young undergraduate population, it is not surprising that the majority exhibited divided and selective attention within normal range. To derive one composite measure of attention, we added each participant's scores on the second and third subtasks. Due to a floor effect resulting from the majority of participants responding within minimum limits of the task, the scores exhibited a strong positive skew. This skew was corrected with an inverse transformation. The mean composite score for the transformed measure was 22.45 (SD = 10.85). After the transformation, higher UFOV scores signified better attentional ability. Although this transformed composite score did not correlate with any other individual difference measures, it did correlate with two outcome measures: mean red zone protection ($r(35) = .40$, $p < .05$) and mean attack efficiency ($r(35) = .33$, $p < .05$) (see Table 2).

Stroop task. Reaction times for correct trials were trimmed to eliminate responses more than three standard deviations above the mean for each condition type for each participant. Trimming did not exceed 3% of trials for any participant. Recall that Kane and Engle (Kane & Engle, 2003) found Stroop task performance differences between low and high span individuals in two measures: facilitation reaction time and interference accuracy. Accordingly, we limited the analysis of this task to those two measures. To arrive at a measure for facilitation reaction time, we subtracted the mean reaction time in the congruent condition from the mean reaction time in the neutral condition for each participant. Recall that facilitation times represent the reduction in response times lower span participants exhibit when they fail to ignore word meaning in the congruent condition, i.e., goal neglect. Greater facilitation reaction time is hypothesized to correlate with lower working memory. Mean facilitation time was 27.0 ms (SD = 41.6). To calculate an interference effect, we subtracted the mean accuracy in the incongruent condition from the mean accuracy in the neutral condition for each participant. The interference effect represents the inability of the participant to resolve the conflict of the mismatched word and hue in the incongruent conditions. Greater interference accuracy is hypothesized to correlate with lower working memory. Mean interference accuracy was 10.2% (SD = 9.5%). Neither facilitation reaction time nor interference accuracy correlated with individual difference or outcome measures. However, facilitation reaction time did correlate with average fixation percentage for the automation interface, $r(35) = -.37$, $p < .05$ (see Table 2).

Simple digit span. Simple digit span reflected the number of correctly recalled digits in all presented sets, with a maximum possible score of 156. Spans for participants ranged from 44 to 148 (M = 92.05, SD = 29.28). In addition to correlating with Aospan measures, participant simple spans also correlated with two outcome measures: mean red zone protection ($r(35) = .44$, $p < .01$) and mean attack efficiency ($r(35) = .34$, $p < .05$) (see Table 2).

Eye tracking measures. To account for differences in eye movement behavior, we calculated the number fixations in the automation messaging interface AOI as percentage of total fixations for each participant. This was performed for each of the four automation scenarios. The average automation fixation percentage was averaged across the four scenarios for each participant to derive one measure. On average, participants rarely fixated on the automation messaging interface AOI (M = 2.79%, SD = 2.71%). To compute average gaze duration, we calculated total gaze time spent in the automation messaging interface AOI divided by number of gazes on the AOI for each of the four automation scenarios. These percentages were averaged across participant to derive one average gaze duration measure. On average, participants spent 141.5 ms (SD = 54.86) on the automation interface each time they gazed upon it. Both eye tracking measures correlated with each other, $r(35) = .50$, $p < .01$. Participants that looked at the automation messaging interface more also looked at it for a longer average duration. Average fixation percentage also correlated with average automation effectiveness, $r(35) = .45$, $p < .01$ (see Table 2).

Table 2. Study One: Correlation Matrix of Individual Difference Measures

	Aospan	UFOV	Facilitation RT	Interference Accuracy	Digit Span	Fixation Percent	Gaze Duration	Red Zone Protection	Attack Efficiency	Automation Effectiveness
Aospan		.15	-.12	-.21	.52**	.08	.00	.24	.20	-.06
UFOV			.04	-.11	.00	-.04	-.29	.40*	.33*	-.16
Facilitation RT				-.10	-.10	-.37*	-.28	.09	.18	-.17
Interference Accuracy					-.09	.15	.14	-.11	-.18	.08
Digit Span						.04	.04	.44**	.34*	-.04
Fixation Percent							.50**	.10	-.07	.45**
Gaze Duration								.16	-.04	.26
Red Zone Protection									.68**	-.25
Attack Efficiency										-.31
Automation Effectiveness										

* $p < .05$. ** $p < .01$.

Outcome Predictors

To determine which individual difference measures were predictive of performance, we regressed the individual difference variables onto each of the performance outcomes. Regression data have been summarized in Table 3.

Red zone protection. Red zone protection scores for each of the four scenarios were averaged across participant. Recall that UFOV performance and digit span positively correlated with red zone protection. These two predictors accounted for 35.0% of the total variance in mean red zone protection, $F(2, 34) = 9.17$, $p < .01$, $R^2_{adjusted} = .31$. Participants with better divided and selective attention ability, as measured via UFOV, exhibited better mean red zone protection, $\beta = .40$, $t(34) = 2.88$, $p < .01$. Likewise, participants with higher simple spans better defended the red zone, $\beta = .44$, $t(34) = 3.16$, $p < .01$.

Attack efficiency. Again, we averaged attack efficiency scores for the four automation scenarios across participants. As with mean red zone protection, UFOV performance ($\beta = .33$, $t(34) = 2.17$, $p < .05$) and digit span positively correlated ($\beta = .34$, $t(34) = 2.23$, $p < .05$) with attack efficiency. These predictors accounted for 22.2% of the total variance in mean attack efficiency, $F(2, 34) = 4.86$, $p < .01$, $R^2_{adjusted} = .18$.

Automation effectiveness. Finally, we averaged automation effectiveness for each of the scenarios to derive one measures for each participant. As noted above, the only individual difference measure that correlated with mean aid effectiveness was average fixation percent allocated to the automation messaging interface. The fixation percent explained 20.6% of the variance in use of the automated aids, $F(1, 35) = 9.06$, $p < .01$,

$R^2_{adjusted}$ = .18. Not surprisingly, the more a participant fixated on the automation messaging interface, the more s/he took advantage of the autonomous aids, $\beta = .45$, $t(35) = 3.01$, $p < .01$.

Table 3. Study One: Multiple Regressions Predicting Performance

Variable	*B*	*SE B*	Β	*t*	$R^2_{adjusted}$
Mean Red Zone Protection					**.31**
UFOV	3.75	1.30	.40	2.88**	
Digit Span	0.00	0.00	.44	3.16**	
Mean Attack Efficiency					**.18**
UFOV	2.75	1.27	.33	2.17*	
Digit Span	0.00	0.00	.34	2.23*	
Mean Aid Effectiveness -					**.18**
Fixation Percent	2.66	0.89	.45	3.01**	

* $p < .05$. ** $p < .01$.

Discussion

The goal of this study was to replicate the findings of the preliminary study with the incorporation of additional individual difference measures and eye tracking. We sought to determine the underlying causes for differences in automation use between high and low span individuals. The task load and level of automation manipulations resulted in many of the same effects on primary performance and automation use seen in the preliminary study. In low task conditions participants exhibited better protection of the red zone, superior attack efficiency, and increased use of the autonomous aids, as compared to high task conditions. Participants also better protected the red zone when they were given one automated aid compared to when they were given two aids. As seen with the preliminary study, the automation once again presented characteristics of "clumsy automation" (Wiener, 1988). Primary performance suffered to a greater extent when high levels of automation were used in high workload conditions, as evidenced by an interaction between these two independent variables on red zone performance. Furthermore, participants used the autonomous aids less in high workload conditions when they were given two aids as compared to one. They also used the autonomous aids to a lesser extent when they were given two aids in the high task load condition compared to the low workload condition. This reiterated the finding that costs of ill-designed automation are greater in high workload conditions.

Unfortunately, the working memory differences observed in the preliminary study did not replicate in study 1. Participant Aospans did not correlate with use of the autonomous aids. This was true for overall mean use of autonomous aids as well as use of autonomous aids in each of the four automation scenarios. The scenarios, Aospan measurement tool, and participant instructions for this study were identical to those of the preliminary study. The lack of correlation can most likely be attributed to sampling error or lack of reliability in the Aospan measurement task. The Aospan task has been shown to consistently predict performance on tasks that require inhibition of distracting information or conflicting goals. It may be possible that the participants of the two studies viewed the nature of the automation task differently. Recall that automation in this paradigm is two-step automation. First, messages from the messaging interface instruct users of the intents of the autonomous aids and 5s later the autonomous aids execute those actions. If the participants of the first study viewed the messaging interface as a distractor, low and high span individuals may have differed in how they responded to it. It is possible that participants in the second study viewed the messaging interface as a secondary extension of the primary task. In this case, both low and high span participants could have shed the messaging task in favor of other strategies to accomplish primary task objectives, thereby eliminating span differences in automation use.

In regards to how Aospan correlated with other individual difference measures, it did correlate with simple digit span, as expected. This finding supported one of our original hypotheses regarding the existence of the phonological loop: working memory is more than just controlled attention and also reflects differences in simple storage. However, Aospan did not correlate with any other individual difference measures. For example, reaction time and accuracy performance on a color Stroop task did not correlate with Aospan, contrary to previous research (Engle, 2002; Kane & Engle, 2003). It is important to note that Kane and Engle (2003) found a strong link between working memory capacity and Stroop performance using an extreme groups design in their analysis. To find differences in performance, they compared the data of individuals with spans in the first and fourth quartiles of Aospan scores. This design may have overestimated the observed effect and may have resulted in an effect that is difficult to replicate (Preacher, Rucker, MacCallum, & Nicewander, 2005; Preacher, 2014). In the Kane and Engle study (2003), the mean OSPAN scores of participants in the first and fourth quartiles were 6.53 (*SD* = 2.05) and 23.25 (*SD* = 6.36), respectively. The mean scores of participants in the first and fourth quartiles of study 1 were 10.67 (*SD* = 6.55) and 62.89 (*SD* = 6.31), respectively. These scores reflect a sample with a higher mean and more overall variability. It is possible that an era of pervasive technology and culture of multitasking created an undergraduate population that scores higher on the Aospan task than seen previously. It is also important to note that although several studies have been published reporting the predictive power of Aospan, studies with null effects are much less likely to be published and are therefore unavailable for comparison.

Despite the lack of correlations among span measures, we did find that some simple measures of individual differences predicted future performance on a complex cognitive task. We found that UFOV task performance and digit span predicted both a participant's ability to defend the red zone and his level of attack efficiency. This finding suggests that participants relied on their ability to divide their visual attention to engage the appropriate enemies. Participants also relied on their simple storage spans to hold relevant information in memory. The predictive nature of simple span is surprising considering that primary performance of the UV tasks did not depend on the incorporation of verbal information. However, some of the information in the messaging interface contained data about the impending location of incoming enemies. Participants with higher simple spans may have been better equipped to use this information to their advantage. Interestingly simple spans did not correlate with use of the autonomous aids, suggesting that participants were reading messages for information about incoming enemies and not reading the messages about the actions of the automated aids. This could have been due to a perceived cost/benefit value of the messages. There was no direct cost for not using the automated aids, but there was a direct cost of a red zone incursion.

The only measure that did predict use of the autonomous aids was the percentage of total fixations that fell on the area of interest defined by the automation messaging interface. In other words, the only predictor of whether or not a participant coordinated his actions with the autonomous aids was if s/he looked at the corresponding message. Fixations on the automation messaging interface indirectly suggest that these participants were better able to divide their attention. However, due to a floor effect in which very few participants looked at the interface, we did not see a correlation between eye tracking data and UFOV. Therefore, to address the second aim of this dissertation, we conducted a follow-up study to design and implement a redesigned automation interface that would encourage looking at the messaging interface, thereby also increasing the use of the automated aids. Because simple span was not associated with use of the autonomous aids (only primary performance measures), a redesign focused on altering the verbal nature of the messages was not considered.

STUDY TWO

Method

Participants

Forty-two George Mason University undergraduate students (24 men and 18 women) with an average age of 21.9 (*SD* = 3.95) years participated in the study for course credit.

Apparatus

The same computer equipment and eye tracker used in study 1 were used for study 2.

Automation Redesign

The same DDD® scenarios used in the first study were also used for study 2. However, a redesigned messaging interface was introduced. The redesign served to mitigate the behavioral limitation observed in study 1: appropriate eye movement behavior. The redesign incorporated an auditory and blinking cue with the presentation of each verbal message to support the central executive by directing attention. Everything else about the messaging interface was identical to the first study (i.e. content of messages, interface location, appearance, etc.).

Procedure

The same individual difference measures used in the first study were collected except for Stroop task performance because it was not associated with any of the outcome variables or individual difference measures. Participants reviewed the same slide presentation about the simulation software and objectives of the scenarios. The presentation was updated to reflect the new automation interface. Participants then completed the same two practice trials administered in the first study to gain familiarity with the simulation and motor controls. Participants then performed each of the four experimental trials in a counterbalanced order. The same eye tracker used in the first study recorded eye movements and the same eye tracking measures were calculated.

Experimental Design

A 2 x 2 repeated-measures factorial design was used to vary task load (low, high) and level of automation (low AA, high AA). In addition to the performance and automation use measures evaluated in the first study, we conducted mixed ANOVAs to determine if the redesigned visual interface influenced primary performance or eye tracking behavior in study 2. We performed the same correlations and regressions to verify if the individual difference and performance measure relationships observed in study 1 persisted in study 2.

Results

Primary Performance Measures

Red zone protection. To assess the effect of task load and level of automation on red zone protection, we conducted a 2 x 2 repeated measures ANOVA. The analysis revealed a main effect of task load $F(1, 41) = 7.17$, $p < .05$, $\eta_p^2 = .149$. Participants more successfully defended the red zone in low task conditions (M = 51.2%, SEM = 2.80%) than in high task load conditions (M = 44.8%, SEM = 2.10%). There was also a main effect for level of automation, $F(1, 41) = 16.34$, $p < .001$, $\eta_p^2 = .285$. Participants exhibited better protection of the red zone with one automated aid (M = 52.2%, SEM = 2.50%) than they did with two aids (M = 43.8%, SEM = 2.30%). Higher level of automation differentially affected red zone protection in the low and high task condition, as evidenced by a significant interaction, $F(1, 41) = 10.01$, $p < .01$, $\eta_p^2 = .196$. In the high task load condition, one aid was helpful (M = 51.6%, SEM = 2.90%) whereas two aids resulted in reduced red zone protection (MD = 38.1%; SEM = 2.00%) (see Figure 6). Both main effects and the interaction were consistent with findings from the preliminary study and study 1.

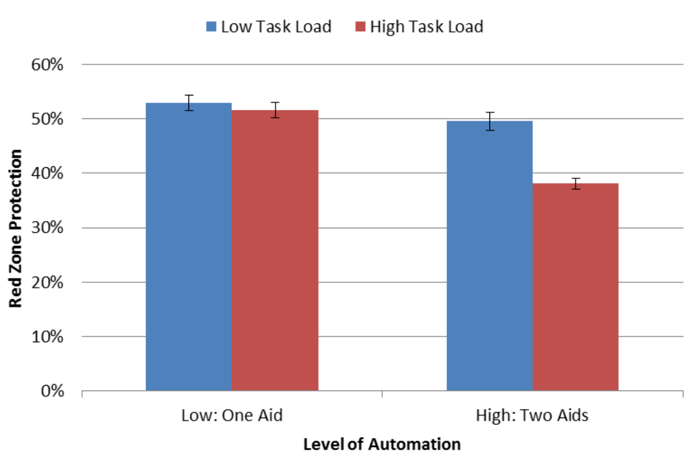

Figure 6. Study Two: Mean red zone protection percentage for low and high task load conditions as a function of level of automation (low, high). Error bars represent standard errors of the mean.

Attack efficiency. To determine the effects of workload and level of automation on attack efficiency, we conducted a 2 x 2 repeated measures ANOVA. Participants exhibited more efficient engagements in low task conditions (M = 72.6%; SEM = 1.90%) compared to high task conditions (M = 69.6%; SEM = 2.00%), as evidenced by a significant main effect of task load $F(1, 41) = 7.94$, $p < .01$, $\boldsymbol{\eta_p^2} = .162$. Level of automation did not influence attack efficiency, $F(1, 41) = 0.04$, $p = .85$. The analysis did not reveal an interaction between level of task load and level of automation, $F(1, 41) = 0.16$, $p = .70$ (see Figure 7). The effects of workload and automation on attack efficiency found in study 1 were replicated in study 2.

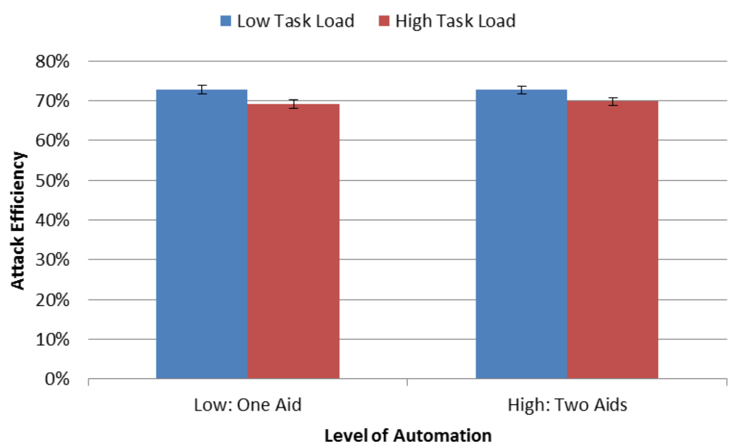

Figure 7. Study Two: Mean attack efficiency for low and high task load conditions as a function of level of automation. Error bars represent standard errors of the mean.

Automation effectiveness. We conducted a final 2 x 2 repeated measures ANOVA on the automation effectiveness measure to determine how level of task load and level of automation influenced participants' use of the automated aids. Level of task load marginally affected use of automation, $F(1, 41) = 2.92$, $p = .10$, $\eta_p^2 = .066$. Contrary to the preliminary study and study 1, in this study participants used the autonomous aids more in high task conditions (M = 60.9%; SEM = 2.40%) than they did in low task load conditions (M = 58.3%; SEM = 2.40%). As with the preliminary study and study 1, level of automation did not influence a participant's use of the autonomous aids, $F(1, 41) = 0.01$, $p = .94$. However, there was a significant interaction between level of task load and level of automation, $F(1, 41) = 8.47$, $p < .01$, $\eta_p^2 = .171$. In the high task conditions participants used the autonomous aids more when they were only given one aid (M = 63.9%; SEM = 2.70%) than they did when they were given two aids (M = 57.9%; SEM = 2.70%) (see Figure 8). This interaction was also found in study 1.

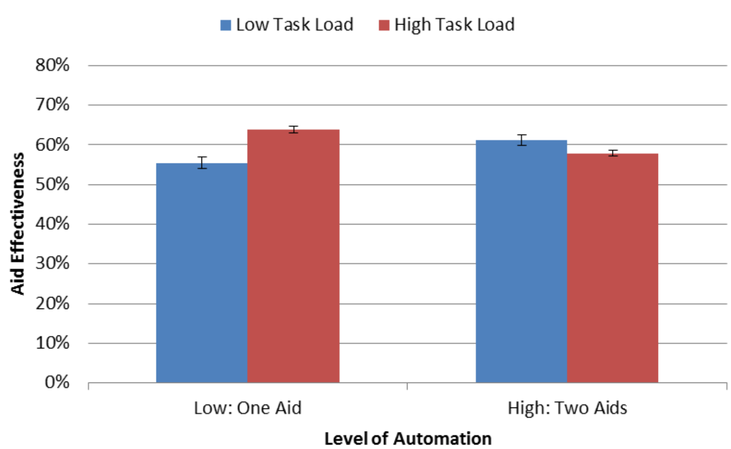

Figure 8. Study Two: Mean use of automation for low and high task load conditions as a function of level of automation (low, high). Error bars represent standard errors of the mean.

Comparisons Across the Two Studies

To contrast primary performance and eye tracking measures between study 1 and study 2, we conducted three-way mixed ANOVAs with study as the between-subjects factor (study 1 vs. study 2) and workload (low vs. high) and level of automation (low vs. high) as the within-subjects factors. Main effects for workload, level of automation, and interactions between workload and level of automation have been discussed above for each study separately and will not be repeated below. Analyses reviewed here address between-subject effects and unique interactions.

Red zone protection. A three-way mixed ANOVA revealed no main effect for study on red zone protection, $F(1, 77) = 1.54$, $p = .21$. Participants exhibited similar levels of red zone protection in study 1 and study 2. Study did not interact with level of workload or level of automation, $F < 3.20$, $p > .07$.

Attack efficiency. There was also no difference between level of attack efficiency between study 1 and study 2, $F(1, 77) = 0.47$, $p = .50$. Study did not interact with level of workload or level of automation, $F < 1.00$, $p > .30$.

Automation effectiveness. The three way mixed ANOVA did reveal a main effect for study, $F(1, 77) = 4.08$, $p < .05$, $\eta_p^2 = .050$. Participants used the autonomous aids 7% less in study 2 (M = 59.6%, SEM = 2.40%) compared to study 1 (M = 66.7%, SEM = 2.50%). Additionally, analyses revealed an interaction between level of workload and study, $F(1, 77) = 7.13$, $p < .01$, $\eta_p^2 = .085$. In study 2 participants used the autonomous aid(s) more in high workload conditions (M = 60.9%, SEM = 2.60%) compared to low workload conditions (M = 58.3%, SEM = 2.40%). Conversely, in study 1 participants used the autonomous aid(s) more in low workload conditions (M = 68.2%, SEM = 2.50%) compared to high workload conditions (M = 65.0%, SEM = 2.70%) (see Figure 9). Study did not interact with level of automation, nor was there a significant three-way interaction between level of workload, level of automation, and study, $F < 1.50$, $p > .25$.

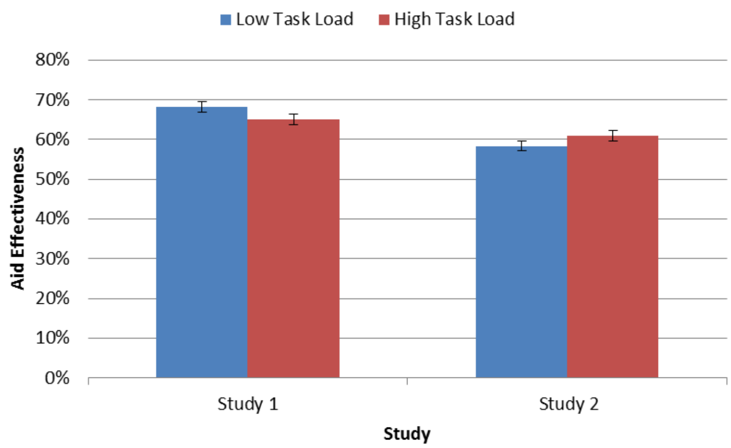

Figure 9. Effect of Study: Mean use of automation for low and high task load conditions as a function of study (study 1, study 2). Error bars represent standard errors of the mean.

Eye movements. A three way mixed ANOVA revealed no difference in the percentage of eye fixations on the messaging interface between study 1 and study 2, $F(1, 69) = 0.002$, $p = .97$. Study did not interact with level of workload or level of automation, nor was there a significant three way interaction, $F < 0.25$, $p > .60$. Study did not affect the duration of the average gaze allocated to the messaging interface, $F(1, 69) = 0.25$, $p = .62$. Study did not interact with level of workload or level of automation, nor was there a significant three way interaction, $F < 2.50$, $p > .15$.

Individual Difference Measures

Aospan. Aospans for participants in study 2 ranged from 6 to 75 (M = 43.88, SD = 17.89). Consistent with study 1, Aospan scores only correlated with simple digit spans, $r(40) = .40$, $p < .01$ (see Table 4).

Useful Field of View (UFOV®). Recall that performance on the second subtask of the UFOV exercise represents divided attention ability. Scores below 100 are considered within normal range. Only one participant fell outside of this limit (M = 19.1, SD = 22.8). The third subtask represents selective attention ability with scores below 350 representing normal performance. All participants' scores for this subtask fell within normal limits (M = 34.7, SD = 36.1). As with study 1, we computed a composite score by combining each participant's performance on subtasks two and three. We then performed an inverse transformation to correct a strong positive skew. The mean score for the transformed composite measure was 25.70 (SD = 9.20). Congruent with findings in study 1, this transformed composite score did not correlate with any other individual difference measures but maintained a correlation with red zone performance, $r(40) = .36$, $p < .05$. However, it no longer correlated with attack efficiency (see Table 4).

Simple digit span. Participant span scores ranged from 44 to 148, M = 91.21, SD = 29.97 (maximum possible score = 156). As expected, digit span scores again correlated with Aospan scores, $r(40) = .40$, $p < .01$. Digit span scores also positively correlated with mean red zone protection ($r(40) = .52$, $p < .001$) and mean attack efficiency ($r(40) = .34$, $p < .05$) (see Table 4).

Eye tracking measures. To assess the amount of visual attention each participant allocated to the messaging interface, we calculated the number of fixations that fell on the interface as a percentage of total recorded fixations for each participant in each scenario. We then averaged across scenarios for each participant to reflect an index of how often each user looked at the messaging interface. Similar to the first study, participants exhibited a low percentage of fixations on the messaging interface, $M = 3.09\%$, $SD = 3.37\%$. To determine how long participants looked at the messaging interface, we computed average gaze duration for each fixation (total time spent on the messaging interface divided by number of gazes) within each scenario. Again, this measure was averaged across scenarios to obtain one measure per user. On average, each time a participant looked at the messaging interface, s/he gazed for 145.9 ms ($SD = 51.53$). The two eye tracking measures correlated with each other; participants that looked at the messaging system more often exhibited longer average gaze durations, $r(40) = .58$, $p < .001$. Finally, the positive relationship between average fixation percentage to the messaging system and automation effectiveness exhibited in study 1 persisted in study 2, $r(40) = .33$, $p < .05$ (see Table 4).

Table 4. Study Two: Correlation Matrix of Individual Difference Measures

	Aospan	UFOV	Digit Span	Fixation Percent	Gaze Duration	Red Zone Protection	Attack Efficiency	Automation Effectiveness
Aospan		.30	.40**	.01	.05	.16	.19	-.14
UFOV			.19	.14	.13	.36*	.16	-.16
Digit Span				.13	.03	.52***	.34*	-.17
Fixation Percent					.58***	.09	.10	.33*
Gaze Duration						.12	.00	.20
Red Zone Protection							.61***	-.31*
Attack Efficiency								-.26
Automation Effectiveness								

* $p < .05$. ** $p < .01$. *** $p < .001$.

Outcome Predictors

To test which individual difference measures predicted performance on the simulation, we repeated the regressions from study 1. Only predictors exhibiting significant correlations with outcomes were included in the analyses (refer to Table 4). Table 5 reflects a summary of the regression models.

Red zone protection. Consistent with findings from study 1, UFOV performance and digit span remained correlated with red zone protection. These two predictors accounted for 34% of the total variance in mean red zone protection, $F(2, 39) = 10.00$, $p < .001$, $R^2_{adjusted} = .31$. Higher levels of controlled and divided attention as measured on the UFOV task corresponded with superior red zone protection, $\beta = .27$, $t(39) = 2.06$, $p <$

.05. Similarly, higher simple digit spans predicted greater red zone protection, $\beta = .46$, $t(39) = 3.50$, $p < .01$.

Attack efficiency. Recall that in study 1 both UFOV performance and simple digit span positively correlated with attack efficiency. In study 2 the relationship between UFOV and attack efficiency failed to reach statistical significance. With digit span as the only predictor, the model accounted for 12% of the total variance in attack efficiency, $F(1, 40) = 5.32$ $p < .05$, $R^2_{adjusted} = .10$. Digit span remained a positive correlate of attack efficiency, $\beta = .34$, $t(40) = 2.31$, $p < .05$.

Automation effectiveness. Finally, fixation percentage remained the only correlate of automation effectiveness. Fixation percent accounted for 11.2% of the total variance in use of the autonomous aids, $F(1, 40) = 5.05$, $p < .05$, $R^2_{adjusted} = .09$. As seen in the previous study, greater fixations allocated to the messaging interface corresponded to greater use of the autonomous aids, $\beta = .34$, $t(40) = 2.25$, $p < .05$. Although red zone protection negatively correlated with automation effectiveness, $r(40) = -.31$, $p < .05$, this measure was not included in the predictive model because it measures primary performance and does not reflect individual differences.

Table 5. Study Two: Multiple Regressions Predicting Performance

Variable	*B*	*SE B*	β	*t*	$R^2_{adjusted}$
Mean Red Zone Protection					**.31**
UFOV	0.004	0.00	.27	2.06*	
Digit Span	0.002	0.00	.46	3.50**	
Mean Attack Efficiency					**.10**
Digit Span	0.00	0.00	.34	2.31*	
Mean Aid Effectiveness -					**.09**
Fixation Percent	1.47	0.66	.34	2.25*	

$^{*}p < .05$. $^{**}p < .01$ $^{***}p < .001$.

Discussion

Based on the finding that eye fixations on the messaging interface predicted use of the autonomous aids in study 1, we designed the second study to increase eye fixations to this interface. We hypothesized that this increase in eye fixations would result in an increase in use of the autonomous aids. To direct more fixations to the interface we incorporated an auditory chime and visual flicker with each incoming message. This redesigned interface did not affect primary performance metrics like red zone protection and attack efficiency. Furthermore, the task load and level of automation effects observed in the preliminary study and study 1 persisted in study 2. Red zone protection and attack efficiency remained higher in low task conditions compared to high task conditions. Participants also better protected the red zone when given one automated aid as opposed to when they were given two. The interaction effect between level of workload and level of automation on red zone protection presented once again. The performance decrement with two aids was larger in the high workload condition compared to the low workload condition.

In regards to use of automation, the interaction effect between level of workload and level of automation on automation use in study 2 replicated the interaction effect seen in the preliminary study and study 1. In low task conditions, participants used the aids more when they were given two aids compared to when they were given one aid. However, in high task conditions participants used the aids more when given only one aid compared to when they were provided two aids. In other words, when participants most needed the aids (under more significant workload) they relied less on the aids meant to assist them.

Although level of workload and level of automation interacted in similar ways to impact automation use in all three studies, level of workload influenced automation use differently in study 2. In the preliminary study and study 1 participants exhibited greater use of the autonomous aids in low workload conditions compared to high workload conditions. However, in study 2 this effect was reversed; participants used the aids more in in high workload conditions compared to low workload (marginal $p = .10$). It is unclear what caused this change in coordination with the autonomous aids. We expected the redesign to result in an increase in use of the autonomous aids across both conditions of task load and were not expecting a differential impact of the redesign. Considering the fact that the addition of the auditory chime and visual flicker to the messaging interface failed to increase eye fixations, it is unlikely to have caused this unusual finding. This reversed impact of workload on automation use could be the result of a random effect. Although the interface redesign was the only difference between study 1 and study 2, there are no theoretical explanations to account for this reversed effect of workload on use of the automated aids.

The redesigned messaging interface did have a significant impact on overall use of the autonomous aids. Contrary to our hypothesis, we discovered that participants used the automated aids 7% *less* in study 2 than they did in study 1. We expected that the addition of the auditory and visual cues to the messaging interface would increase fixations to the interface and promote a subsequent increase in participant coordination with the autonomous aids. However, the redesign did not affect fixation percentage on the interface. In both studies participants spent three percent of total fixations on the area of interest containing the messages. Despite allocating an equal proportion of eye fixations to the messaging system, participants in study 2 exhibited greater disuse of the autonomous aids. One possible explanation for this reduction in automation use is that the introduction of the chime and flicker to the messaging system could have distracted participants and reduced their ability to coordinate efforts with the aids. It is possible that the additional auditory and visual activity added to the cognitive load of participants without changing overall eye movement behaviors. Further investigation of the redesign is required to test this hypothesis. However, obtaining an empirical measure of cognitive load/distraction would require the collection and analysis of physiological measures to avoid the introduction of subjective surveys or intrusive secondary tasks. Using physiological measures as an index of cognitive load is a laborious process riddled with its own set of limitations due to individual differences.

Another explanation for the decreased use of the autonomous aids in study 2 relates to the unexpected negative correlation between red zone protection and automation effectiveness. This negative relationship describes operators that successfully defended the red zone without needing to coordinate their efforts with the autonomous aids. Because there was no negative cost to not coordinating with the aids, these operators most likely made a conscious decision to disuse the aids and engaged more enemies on their own than was necessary. Although we designed the tasks to reflect high task load to encourage automation use, it may not have been difficult enough for these over achievers. Further research should investigate if these high performers possess high or low working memory spans and examine their data separately from users who disuse automation because of inefficient attention allocation or work overload.

Although the redesign impacted the nature of automation use, it did not affect the predictive nature of the individual difference measures. Once again, participant Aospan scores did not correlate with any primary performance measures or use of automation. As with study 1, working memory scores only correlated with simple digit scores. More interestingly, study 2 replicated the finding that simple digit spans positively correlated with and predicted red zone protection and attack efficiency. Controlled attention as measured via UFOV remained correlated with red zone performance in study 2, but was no longer correlated to attack efficiency. These relationships suggest that primary performance on complex command and control tasks requires both simple short term memory and controlled attention. Measuring these individual differences provide predictive value for future tasks and merit further investigation. The practical implications of these results should also be explored to determine if training short term memory and controlled attention can improve command and control performance. However, in terms of use of automation the only predictor in both studies 1 and 2 remained average fixation percentage on the messaging interface. The individual differences governing automation use/misuse are largely dependent on the specific context of the task are more complex that can be fully understood by this series of experiments.

CONCLUSION

Through these experiments we sought to: 1) better understand how differences in working memory capacity influence human-automation interactions in a simulated air defense task and 2) investigate if automation redesign can mitigate differences in automation misuse caused by working memory deficiencies.

In regards to the first goal, unfortunately these studies were inconclusive in explaining the relationship between working memory capacity and automation use. In the preliminary study differences in working memory capacity correlated with automation use. However, this correlation did not replicate in studies 1 or 2. Nevertheless, we observed two notable related findings that merit further investigation. Firstly, although WMC did not correlate directly with primary performance or automation use, we found that differences in *short term memory* provided predictive value for primary performance. In other words, differences in phonological loop capabilities also influenced performance on complex cognitive tasks, not just controlled attention. Differences in short term memory and their contribution to higher level tasks have been discounted by Engle (2002) and Kane et al. (2001). Yet, in both studies 1 and 2 simple digit spans explained additional variance in red zone protection and attack efficiency above and beyond the variance that controlled attention (via UFOV) explained alone. Future studies should investigate if training short term memory can improve performance of command and control tasks. If this relationship between short term memory and primary performance is replicated and better understood, practical applications related to personnel selection and training should also be explored.

Secondly, these experiments revealed that we could use eye tracking measures to predict automation use. Despite the fact that WMC did not predict automation use, eye fixation percentage on the messaging system consistently correlated with automation use in both studies 1 and 2. In fact, eye fixation percentage explained between 11 and 21 percent of the variance in automation use. This provides evidence for real-time unobtrusive eye measurements as a means to monitor automation use. System designers can implement interventions to maintain automation use within predetermined ranges by using eye movements as a more continuous measure of automation use. This measure could be particularly helpful in automation that is communicated and implemented in multiple stages with varying levels. For example if an automatic braking system needs prior user approval or operator input before being executed, eye tracking measures to the first stage of the automation can alert potential disuse in the second stage of execution. Although our redesigned interface did not successfully increase eye movements, other redesigns may prove more successful.

In regards to exploring if automation redesign can mitigate automation disuse, again our findings were inconclusive. However, an important lesson was learned. We discovered that redesigning an automation interface does not always produce the outcomes expected. In fact, we saw that our redesign resulted in unanticipated changes in user behavior (Parasuraman & Riley, 1997). In study 1 we found that eye fixation percentage on the messaging interface was the only predictor of automation use. Therefore, in study 2 we redesigned the messaging interface to increase these fixations by adding an auditory chime and visual flicker to the interface. In simple tasks the addition of an auditory and visual cue would undoubtedly result in more fixations to the cue. Thus, we hoped this would encourage increased fixations to the messaging interface and subsequently increase use of the autonomous aids. However, the addition of an auditory chime and visual flicker to the messaging system did not increase eye fixations to the interface and, in fact, decreased use of the automated aids. We can only speculate as to why this happened. One theory to explain this reduction in automation use is that the redesign may have added cognitive load to the participants by introducing a visual and auditory distraction. This theory requires further investigation, but provides an example of Parasuraman and Riley's (1997) warning that automation (and its redesigns) can change the way human interact with systems in unintended ways.

Limitations and Future Research

Overall the findings from these studies reveal the multifaceted relationship between WMC, complex decision-making processes like command and control tasks, and automation use. These relationships are not as straight forward as the stable correlations found between WMC and other cognitive tasks (e.g. reading comprehension). Many potential factors must be accounted for to fully understand how WMC contributes to automation use. Some of these factors include: trust in the automation; the user's confidence in his/her abilities compared to the automation; and abilities and skills of the user such as spatial span, mental rotation, and video game experience. These factors are complicated by the fact that many of these variables interact with each other. For example, some participants exhibited very high red zone protection (primary task performance) and high WMC with little to no use of the autonomous aids (as see in study 2). In this example, misuse of automation may have been the result of a conscious strategy and less related to WMC. Because there was no direct negative cost of misuse of automation in our paradigm, this strategy was not necessarily a bad one. Therefore, future studies should incorporate user interviews to understand the motivations behind automation use and design the automation so there is a more dependent reliance on automation use and successful primary task performance (e.g. incorporate a cost of disuse). Recall that operator skill on the primary task predicted use of automation in both the preliminary study and study 2. Therefore, skill level should be controlled for when exploring individual difference measures and how they predict automation use. Furthermore, other individual difference measures such as spatial span and video game experience should be collected to investigate other predictors of automation use.

Despite these limitations and unexpected findings, this series of studies reinforces the importance of individual differences, including short term memory, on primary task performance, provides evidence for eye movements as a continuous measure of automation use, and serves as an example of how automation design (and redesign) can result in unintended user behavior.

REFERENCES

Ackerman, P. L., Beier, M. E., & Boyle, M. D. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General*, *131*(4), 567–589. doi:10.1037//0096-3445.131.4.567

Ahmed, N., de Visser, E., Shaw, T., Mohamed-Ameen, A., Campbell, M., & Parasuraman, R. (2014). Statistical modelling of networked human-automation performance using working memory capacity. *Ergonomics*, *57*(3), 295–318. doi:10.1080/00140139.2013.855823

Arthur, W. J., Bennett, W. J., Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance, 11*(1), 57–101. doi:10.1207/s15327043hup1101_3

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423. doi:10.1016/S1364-6613(00)01538-2

Baddeley, A. (2002). Is working memory still working? *European Psychologist*, *7*(2), 85–97. doi:10.1027//1016-9040.7.2.85

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829–39. doi:10.1038/nrn1201

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1–29. doi:10.1146/annurev-psych-120710-100422

Baddeley, A., & Hitch, G. J. (1974). Working Memory. In *The psychology of learning and motivation* (Vol. 8, pp. 47–89).

Bahner, J. E., Hüper, A.-D., & Manzey, D. H. (2008). Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *International Journal of Human-Computer Studies, 66*(9), 688–699. doi:10.1016/j.ijhcs.2008.06.001

Ball, K., Beard, B., Roenker, D., Miller, R., & Griggs, D. (1988). Age and visual search: Expanding the useful field of view. *JOSA A*, *5*(12), 2210–2219.

Ball, K., & Owsley, C. (1993). The useful field of view test: a new technique for evaluating age-related declines in visual function. *Journal of the American Optometric Association*, *64*(1), 71–79.

Ball, K., Owsley, C., Sloane, M. E., Roenker, D. L., & Bruni, J. R. (1993). Visual attention problems as a predictor of vehicle crashes in older drivers. *Investigative Ophthalmology & Visual Science*, *34*(11), 3110–23.

Barnes, M., & Grossman, J. (1985). *The intelligent assistant concept for electronic warfare systems*. *Tech. Rep. NWC TP, 5585*. China Lake, CA,.

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General*, *132*(1), 71–92.

Bleckley, M. K., Durso, F. T., Crutchfield, J. M., Engle, R. W., & Khanna, M. M. (2003). Individual differences in working memory capacity predict visual attention allocation. *Psychonomic Bulletin & Review*, *10*(4), 884–9.

Broadbent, D. (1958). *Perception and Communications*. New York, NY: Permagon Press.

Bühner, M., König, C. J., Pick, M., & Krumm, S. (2006). Working memory dimensions as differential predictors of the speed and error aspect of multitasking performance. *Human Performance*, *19*(3), 253–275. doi:10.1207/s15327043hup1903_4

Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology*, *42*(3), 249–268.

Chen, J. Y. C., & Barnes, M. (2012). Supervisory control of multiple robots: Effects of imperfect automation and individual differences. *Human Factors*, *54*(2), 157–174. doi:10.1177/0018720811435843

Chen, J. Y. C., & Terrence, P. I. (2009). Effects of imperfect automation and individual differences on concurrent performance of military and robotics tasks in a simulated multitasking environment. *Ergonomics*, *52*(8), 907–920. doi:10.1080/00140130802680773

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. doi:10.1016/S0022-5371(80)90312-6

De Visser, E., Shaw, T., Mohamed-Ameen, A., & Parasuraman, R. (2010). Modeling human-automation team performance in networked systems: Individual differences in working memory count. In *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting* (Vol. 54, pp. 1087–1091). San Francisco, CA: HFES.

De Winter, J. C. F., & Dodou, D. (2014). Why the Fitts list has persisted throughout the history of function allocation. *Cognition, Technology & Work*, *16*(1), 1–11. doi:10.1007/s10111-011-0188-1

Endsley, M. R., & Kiris, E. O. (1995). The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Hu*, *37*(2), 381–394. doi:10.1518/001872095779064555

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19–23.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 44, pp. 145–199). New York, NY: Elsevier.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp. 102–134).

Feigh, K. M., Dorneich, M. C., & Hayes, C. C. (2012). Toward a Characterization of Adaptive Systems: A Framework for Researchers and System Designers. *Hu*, *54*(6), 1008–1024. doi:10.1177/0018720812443983

Fitts, P., Jones, R., & Milton, J. (1950). Eye movements of aircraft pilots during instrument-landing approaches. *Aeronautical Engineering Review*, *9*(2), 24–29.

Grabowski, M., & Hendrick, H. (1993). How low can we go?: Validation and verification of a decision support system for safe shipboard manning. *Engineering Management, IEEE Transactions on*, *40*(1).

Inagaki, T. (2003). Adaptive automation: Sharing and trading of control. In E. Hollnagel (Ed.), *Handbook of cognitive task design* (pp. 147–169). Mahwah, New Jersey: Erlbaum.

Jarrold, C., & Towse, J. N. (2006). Individual differences in working memory. *Neuroscience*, *139*(1), 39–50. doi:10.1016/j.neuroscience.2005.07.002

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. doi:10.1080/1463922021000054335

Kaber, D. B., Omal, E., & Endsley, M. R. (1999). Level of Automation Effects on Telerobot Performance and Human Operator Situation Awareness and Subjective Workload. In M. W. Scerbo & M. Mouloua (Eds.), *Automation technology and human performance: Current research and trends* (pp. 165–170). Mahwah, New Jersey: Lawrence Erlbaum.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*(2), 169–183.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*(1), 47–70. doi:10.1037/0096-3445.132.1.47

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189–217.

Kirlik, A. (1993). Modeling strategic behavior in human-automation interaction: Why an "aid" can (and should) go unused. *Hu*, *35*(2), 221–242. doi:10.1177/001872089303500203

Lathan, C. E., & Tracey, M. (2002). The effects of operator spatial perception and sensory feedback on human-robot teleoperation performance. *Presence: Teleoperators and Virtual Environments*, *11*(4), 368–377. doi:10.1162/105474602760204282

Lee, J. D., & Seppelt, B. D. (2009). Human factors in automation design. In S. Y. Nof (Ed.), *Springer handbook of automation* (pp. 417–436). Berlin, Germany: Springer Berlin Heidelberg. doi:10.1007/978-3-540-78831-7_25

McKendrick, R., Shaw, T., De Visser, E., Saqer, H., Kidwell, B., & Parasuraman, R. (2013). Team Performance in Networked Supervisory Control of Unmanned Air Vehicles: Effects of Automation, Working Memory, and Communication Content. *Hu*. doi:10.1177/0018720813496269

Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors*, *50*(2), 194–210. doi:10.1518/001872008X288574

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(1), 81–97.

Mueller, S. T. (2012). The Psychology Experiment Building Language, Version 0.13.

Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity — facets of a cognitive ability construct. *Personality and Individual Differences*, *29*(6), 1017–1045. doi:10.1016/S0191-8869(99)00251-2

Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2013). Human Performance Consequences of Stages and Levels of Automation: An Integrated Meta-Analysis. *Hu*, *56*(3), 476–488. doi:10.1177/0018720813501549

Owsley, C., Ball, K., McGwin, G., Sloane, M. E., Roenker, D. L., White, M. F., & Overley, E. T. (1998). Visual processing impairment and risk of motor vehicle crash among older adults. *JAMA : The Journal of the American Medical Association*, *279*(14), 1083–8.

Owsley, C., Mcgwin Jr, G., & Ball, K. (1998). Vision impairment, eye disease, and injurious motor vehicle crashes in the elderly. *Ophthalmic Epidemiology*, *5*(2), 101–114.

Parasuraman, R. (2000). Designing automation for human use: empirical studies and quantitative models. *Ergonomics*, *43*(7), 931–951. doi:10.1080/001401300409125

Parasuraman, R., Bahri, T., Deaton, J. E., Morrison, J. G., & Barnes, M. (1992). *Theory and design of adaptive automation in aviation systems. (Tech Report NAWCADWAR-92033-60).* Warminster, PA: Naval Air Warfar Center.

Parasuraman, R., Cosenzo, K., & de Visser, E. (2009). Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. *Military Psychology*, *21*(2), 270–297. doi:10.1080/08995600902768800

Parasuraman, R., Kidwell, B., Olmstead, R., Lin, M.-K., Jankord, R., & Greenwood, P. (2013). Interactive effects of the COMT gene and training on individual differences in supervisory control of unmanned vehicles. *Human Factors*. doi:10.1177/0018720813510736

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *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. doi:10.1207/s15327108ijap0301_1

Parasuraman, R., Mouloua, M., & Hilburn, B. (1999). Adaptive aiding and adaptive task allocation enhance human-machine interaction. In M. W. Scerbo & M. Mouloua (Eds.), *Automation technology and human performance: Current research and trends* (pp. 119–123). Mahwah, New Jersey: Lawrence Erlbaum.

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors*, *38*(4), 665–679.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230–253. doi:10.1518/001872097778543886

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *30*(3), 286–297.

Preacher, K. J. (2014). The Encyclopedia of Clinical Psychology, *2*, 1189–1192. doi:10.1002/9781118625392.wbecp190

Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: a critical reexamination and new recommendations. *Psychological Methods*, *10*(2), 178–92. doi:10.1037/1082-989X.10.2.178

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., & Pope, A. T. (2003). Effects of a psychophysiological system for adaptive automation on performance, workload, and the event-related potential P300 component. *Human Factors, 45*(4), 601–613. doi:10.1518/hfes.45.4.601.27092

Repovs, G., & Baddeley, A. (2006). The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience*, *139*(1), 5–21. doi:10.1016/j.neuroscience.2005.12.061

Sanders, A. (1970). Some aspects of the selective process in the functional visual field. *Ergonomics*, *13*(1), 101–117.

Saqer, H., & Parasuraman, R. (2014). Individual performance markers and working memory predict supervisory control proficiency and effective use of adaptive automation. *International Journal of Human Factors and Ergonomics*, *3*(1), 15–31.

Sarter, N. B., Woods, D. D., & Billings, C. (1997). Automation surprises. *Handbook of Human Factors ...*, 1926–1943.

Scerbo, M. W. (2007). Adaptive automation. In R. Parasuraman & M. Rizzo (Eds.), *Neuroergonomics: The brain at work* (pp. 239–252). New York, NY: Oxford University Press.

Scerbo, M. W., Freeman, F. G., Mikulka, P. J., Parasuraman, R., Di Nocera, F., & Prinzel, L. J. (2001). *The efficacy of psychophysiological measures for implementing adaptive technology*. *(Tech Report NASA TP-2001-211018)*. Hampton, VA: NASA Langley Research Center.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, *125*(1), 4–27. doi:10.1037/0096-3445.125.1.4

Sheridan, T. B. (2011). Adaptive automation, level of Automation, allocation authority, supervisory control, and adaptive control: Distinctions and modes of adaptation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *41*(4), 662–667. doi:10.1109/TSMCA.2010.2093888

Singh, I. L., Molloy, R., & Parasuraman, R. (1993a). Automation-induced "Complacency": Development of the complacency-potential rating scale. *The International Journal of Aviation Psychology*, *3*(2), 111–122. doi:10.1207/s15327108ijap0302_2

Singh, I. L., Molloy, R., & Parasuraman, R. (1993b). Individual differences in monitoring failures of automation. *Journal of General Psychology, 120*(3), 357–373.

Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence*, *30*(3), 261–288. doi:10.1016/S0160-2896(01)00100-3

Syrdal, D. S., Lee Koay, K., Walters, M. L., & Dautenhahn, K. (2007). A personalized robot companion?-The role of individual differences on spatial preferences in HRI scenarios. In *The 16th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 1143–1148). Jeju, Korea: IEEE. doi:10.1109/ROMAN.2007.4415252

Szalma, J. L., & Taylor, G. S. (2011). Individual differences in response to automation: the five factor model of personality. *Journal of Experimental Psychology: Applied*, *17*(2), 71–96. doi:10.1037/a0024170

Turner, M., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127–154. doi:10.1016/0749-596X(89)90040-5

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498–505.

Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1302–21. doi:10.1037/0278-7393.30.6.1302

Wiener, E. L. (1988). Cockpit automation. In E. L. Wiener & D. C. Nagel (Eds.), *Human factors in aviation* (pp. 433–461). San Diego, CA: Academic Press.

Wilson, G. F., & Russell, C. A. (2007). Performance Enhancement in an Uninhabited Air Vehicle Task Using Psychophysiologically Determined Adaptive Aiding. *Human Factors*, *49*(6), 1005–1018. doi:10.1518/001872007X249875

BIOGRAPHY

Haneen Rezik Saqer received her Bachelor of Arts in Psychology from Trinity University, San Antonio, Texas in 2001 after which she received a Master of Health Administration from Washington University School of Medicine in St. Louis, Missouri in 2003. She enjoyed a career as a medical business consultant and practice administrator until 2009 when she enrolled in the doctoral program in Human Factors and Applied Cognition at George Mason University under the advisement of Dr. Raja Parasuraman. She received her Master of Arts in Psychology with a concentration in Human Factors and Applied Cognition in 2011. During her time as a graduate student her research focused on individual differences in human-automation interaction, decision-making and team dynamics in human-robot interaction, and the use of eye movements as an online measure of situation awareness. She also served as a teaching assistant for six sections of Statistics in Psychology, four sections of Introduction to Psychology, and two sections of Sensation and Perception. Between school years she gained valuable industry experience through three internships at PROS, Google [x], and Dell Inc. Haneen was also a founding member of a driver distraction prevention program aimed at teenage drivers, "Do Not Disturb."