7-2016

The effect of individual differences in working memory in older adults on performance with different degrees of automated technology

Richard Pak
Clemson University, richpak@clemson.edu

Anne Collins McLaughlin
North Carolina State University at Raleigh

William Leidheiser
Clemson University

Ericka Rovira
U.S. Military Academy Department of Behavioral Sciences & Leadership, West Point, NY

The effect of individual differences in working memory in older adults on performance with different degrees of automated technology

Richard Pak[1], Anne Collins McLaughlin[2], William Leidheiser[1] , & Ericka Rovira[3]

[1]*Clemson University Department of Psychology, Clemson, SC 29634*

[2]*North Carolina State University Department of Psychology, Raleigh, NC 27695*

[3]*U.S. Military Academy Department of Behavioral Sciences & Leadership, West Point, NY 10996*

Corresponding author:

Richard Pak

Clemson University

Department of Psychology

418 Brackett Hall

Clemson, SC 29634

USA

richpak@clemson.edu

The effect of individual differences in working memory in older adults on performance with different degrees of automated technology

A leading hypothesis to explain older adults' over-dependence on automation is age-related declines in working memory. However, it has not been empirically examined. The purpose of the current experiment was to examine how working memory affected performance with different degrees of automation in older adults. In contrast to the well-supported idea that higher degrees of automation, when the automation is correct, benefits performance but higher degrees of automation, when the automation fails, increasingly harms performance, older adults benefited from higher degrees of automation when the automation was correct but were not differentially harmed by automation failures. Surprisingly, working memory did not interact with degree of automation but did interact with automation correctness or failure. When automation was correct, older adults with higher working memory ability had better performance than those with lower abilities. But when automation was incorrect, all older adults, regardless of working memory ability, performed poorly.

Practitioners summary: The design of automation intended for older adults should focus on ways of making the correctness of the automation apparent to the older user and suggest ways of helping them recover when it is malfunctioning.

Keywords: working memory, degree of automation, older adults, trust, complacency, automation bias

1.0 Introduction

As automated technology appears in more work and home settings, it is important to understand how all types of users perform with automation, however previous research has tended to focus on younger users. The cognitive support provided by automated technologies may be even more important for older adults. Contrary to commonly held stereotypes, older adults are very open to technological assistance (Mitzner et al., 2010). For example, automated technologies can support independent mobility (personal GPS devices, autonomous vehicles), health (blood glucose meters), and finances (online banking). Automated technologies may be key to helping older adults maintain an independent (i.e., outside of an assisted living facility) and productive lifestyle.

Although the study of age-related differences in performance with automated technology has been sparse (see Hoff & Bashir, 2015 for a brief review), one of the most consistent findings in this literature is that older adults tend to over-depend and over-trust automation compared to other age groups (Donmez, Boyle, Lee, McGehee, 2006; Fox & Boehm-Davis, 1998; Ho, Wheatley, & Scialfa, 2005; Lee et al, 1997; Pak et al., 2012, 2014). An illustrative example is that compared to younger adults, older adults were more likely to trust in-vehicle distraction mitigation technology even when it was demonstrated to be incorrect (Donmez, Boyle, Lee, & McGehee, 2006). This could be seen as yet another example of the paradoxical observation in aging research that show sharp inconsistencies between older adults' beliefs about their ability and actual performance (de Winter, Dodou, & Hancock, 2015). This paradoxical behaviour with automation presents a possibly dangerous situation (e.g., Hancock, 2014) especially if older adults, who could benefit most from the use of automation to maintain an independent and engaged life, end up over-depending on imperfect automation.

1.1 Older Adults and Automation

A leading hypothesis to explain older adults' characteristic performance with automation is age-related decline in working memory (Ho, Kiff, Plocher, & Haigh, 2005; Ho, Wheatley, & Scialfa, 2005). Working memory is the ability to maintain information over a short period despite distraction, but also the capacity for controlled processing (Engle, Tuholski, Laughlin, & Conway, 1999; Unsworth & Engle, 2007). In one study, Ho, Wheatley, and Scialfa had younger and older adults take appropriate medications at the right time and dose using an imperfectly automated medication management (AMM) system. When it was time to take one of four medications, the AMM displayed a message indicating medication and dosage. Compared to younger adults, older adults were more compliant with (uncritically agreeing) and reliant on (not verifying) the AMM. The authors explained that this pattern was likely due to age-related differences in working memory. Their reasoning was that younger adults with presumably higher working memory abilities were likely using a strategy where they simply remembered the medication regimen and thus did not need to use the AMM; a strategy older adults could not use given their reduced working memory abilities. Lee and Moray (1994) found that when self-confidence in a task is lower than their perceived trust, participants tended to depend on automation. Thus, older adults' awareness of their declining memory abilities may have contributed to a sharply reduced self-confidence in their ability to carry out the task compared to their level of trust in automation leading to their higher dependence on the AMM.

Ho, Kiff, Plocher, and Haigh (2005) theorised another possible way in which age deficits in working memory might affect automation behaviour. The acquisition of mental models is known to be working memory dependent (Gilbert & Rogers, 1999); thus age-related declines in working memory may result in incomplete or inaccurate representations of the task leading them to use automation more frequently. However, it is possible that the same working memory declines may cause older adults to develop inaccurate mental representations of automation itself leading to poorer performance with automation (Kieras & Bovair, 1984). Sheridan (1992) referred to this as a user's ability to understand automation and suggested that accurate mental representations of automation are important to performance (Sarter, Mumaw, & Wickens, 2007). The relationship between older adults' less accurate mental representation of automation and poorer performance was also found in a study by Olson, Fisk, and Rogers (2011).

Older adults' increased tendency to comply with and rely on automation even when that automation is incorrect was also observed in a study by McBride, Fisk, and Rogers (2011). The goal of their study was to examine the effect of workload and age on compliance with and reliance on an automated warehouse management system (AWMS). Compliance is a behaviour where the operator adheres to unreliable automation while reliance is where the operator *does not act* (such as not verifying) when the automation fails to alert (Dixon, Wickens, & McCarley, 2007). They found that, while increased workload affected younger adult's compliance with and reliance on automation, older adults were unaffected by workload but had generally high compliance and reliance rates compared to younger adults. Because the focus of their study was on the effect of workload on age-related differences in automation compliance and reliance, they did not attempt to explain the source of older adults generally higher levels of compliance and reliance with automation.

What is noteworthy about Ho, Wheatley, and Scialfa's (2005) and McBride, Fisk, and Rogers' (2011) studies was that both systems were examples of alerting or warning automation which represent a lower degree of automation support (Onnasch, Wickens, Li, & Manzey, 2014). Degree of automation is a concept that describes what and how much a task is being automated (Onnasch, Wickens, Li, & Manzey; Wickens, Li, Santamaria, Sebok, & Sarter, 2010). It is a more convenient way to express what is being automated (stages) and by how much (levels). Higher degrees could indicate both increased stage of automation or increased level of automation (Wickens, Li, Santamaria, Sebok, & Sarter, 2010). The purpose of alerting automation is to monitor the environment to gather and analyse information and inform the user so they can shift attention, but it is the user's responsibility to interpret the automation and make a decision. For example, the AMM used by Ho, Wheatley, and Scialfa (2005), monitored time to take medication while the AWMS used by McBride, Fisk, and Rogers (2011) alerted the user to the presence of specific packages or filled trucks in the warehouse. These kinds of automated systems are thought to rely on attentional abilities and attentional strategies, particularly when reliability is low (Onnasch, Ruff, & Manzey, 2014). But it is interesting to note that while Ho, Wheatley, and Scialfa (2005) found that older adults increased their monitoring, it did not improve performance. For this reason, they ruled out age-differences in attention as an explanation for age-related differences in automation behaviour and hypothesised age-related differences in working memory as the reason for older adult's poorer performance.

Alerting automation is contrasted with decision aids, a higher degree of automation, that assists a user in making a decision in an uncertain or time-constrained situation (Bahner, Hüper, & Manzey, 2008; Skitka, Mosier, & Burdick, 1999). Compared to alerting automation, decision aids carry out relatively more of the task while the user does less (Parasuraman, Sheridan, & Wickens, 2000). While the literature on aging and automation has mainly focused on alerting automation, Pak et al (2012) examined older adults using decision aids and found that older adults also had generally high levels of trust compared to younger adults. However, the focus of their study was trust and the design of automation so they did not examine the relationship of working memory or other abilities on older adults trust and dependence on automation.

1.2 Working Memory and Degrees of Automation

Studies with younger adults have established the importance of working memory on how users respond to automated technologies, albeit indirectly (de Visser, Shaw, Mohamed-Ameen, and Parasuraman, 2010; Saqer, de Visser, Emfield & Parasuraman, 2011; Parasuraman, de Visser, Lin, & Greenwood, 2012; McKendrick, Shaw, de Visser, Saqer, Kidwell, & Parasuraman, 2013; Ahmed, de Visser, Shaw, Mohamed-Ameen, Campbell, & Parasuraman, 2014; [blinded], in review). Neuro-genetic evidence from Parasuraman, de Visser, Lin, and Greenwood showed that individuals who possessed genes that appear to be related to lower working memory, showed greater negative performance effects when the automation failed than the group possessing genes related to higher working memory. The authors theorised that low working memory individuals have a reduced ability to update the contents of working memory which reduces the amount of time and resources available to consider and confirm an automation failure. However, the results of the study only indirectly implicated effects of working memory on automation performance because working memory ability scores were not obtained and the authors stress the need to directly examine the role of working memory on the appropriate use of automation.

More direct evidence of the importance of working memory was found in a recent study by [blinded] (in review). The researchers wanted to examine the purported link between working memory but to also extend the research to examine different degrees of automation. The link between working memory and degrees of automation was theorised to come from the fact that different degrees of automation assume different kinds of cognitive tasks from the user (Parasuraman, Sheridan, & Wickens, 2000); higher degrees of automation support the operator in decision-making which is working memory-intensive because operators need to generate and compare different courses of action (e.g., GPS guided route guidance) whereas lower degrees of automation support the operator by taking away some of the information acquisition and analysis properties of the task (e.g., vehicle check engine light notifies the operator of a problem, however the user must diagnose the issue). They found a critical interaction such that when automation degree was higher (calculating distances, narrowing choices, and assigning priority for the user) low working memory individuals' performance was both more helped by automation that was correct and more harmed by automation failures than high working memory individuals. Their study provided the first empirical demonstration of the link between working memory capacity and behaviour with automation of different degrees, however, this study was done only with younger adults.

1.3 Knowledge Gaps in the Study of Aging and Automation

Age-related declines in working memory are some of the most robust findings in cognitive aging (e.g., Salthouse & Babcock, 1991, Verhaeghen & Salthouse, 1997; Salthouse, 1992, 1996). However, the previously discussed studies examining working memory and automated technologies have used a younger adult sample (e.g., de Visser, Shaw, Mohamed-Ameen, & Parasuraman, 2010; Parasuraman, de Visser, Lin, and Greenwood, 2012; Saqer, de Visser, Emfield & Parasuraman, 2011; [blinded], in review). Ho, Wheatley, and Scialfa's (2005) study that originally hypothesised the critical role of age-related differences in working memory as an explanation for age-related differences with automation did not directly measure working memory span but merely inferred its role based on established literature on age differences in working memory.

Another limitation of existing studies (e.g., Donmez, Boyle, Lee, & McGehee, 2006; Ezer, Fisk, & Rogers, 2008; Ho, Wheatley, Scialfa, 2005; McBride, Fisk, & Rogers, 2011; Sanchez, Rogers, Fisk, & Rovira, 2012; Wiegmann, McCarley, Kramer, & Wickens, 2006) is that they have all mostly focused on simpler, binary-state alerting automation and have not examined the extent to which various degrees of automation affect performance in older adults. The distinction between lower degree alerting automation and higher degree decision automation is of major interest because there are distinct human performance consequences with different degrees of automation (Crocoll & Coury, 1990; Onnasch, Wickens, Li, & Manzey, 2014; Parasuraman & Manzey, 2010; Rovira, et al. 2007), especially with regard to working memory ability ([blinded], in review). However, no extant research has examined how older adult's performance is affected by moving across this boundary.

2.0 Overview of the Experiment

The purpose of the current experiment was to address two significant gaps in the research on older adults and automation: 1) how do individual differences in working memory in older adults affect automation performance, and 2) extend the literature by examining how older adults working memory abilities affect performance with different degrees of automation.

Given the sparse amount of research on abilities, automation, and aging, our hypotheses are generated by extrapolating findings from extant younger adult studies and the wider cognitive aging literature. The most direct comparison can be made with [blinded]'s (in review) study examining the performance of low and high working memory span in younger adults.

(1) Consistent with previous literature, overall accuracy will be better when the automation is correct compared to trials where the automation fails.

(2) Older adults will also show higher accuracy with reliable decision automation (a higher degree of automation) compared to information automation (a lower degree), reflecting the beneficial performance effects of higher, reliable automation, (Onnasch et al., 2014).

    (a) This benefit of higher degrees of reliable automation will be moderated by individual differences in working memory abilities.

(3) But when the automation fails, performance will be worse with decision compared to information automation, reflecting the consequences of not being

involved in the decision-making aspects of the task and thus being out of the loop (Wickens, 1992).

(a) The magnitude of this performance penalty of failures of the decision automation will be related to individual differences in working memory abilities.

(4) Given prior research that has shown older adults' generally increased complacency and levels of trust in automation, we expect that measures of trust will be strongly correlated to performance and reflective of complacency, indicated as high levels of trust in automation regardless of its correctness.

2.1 Method

2.1.1 Participants

Forty-four community-dwelling older adults (24 females; mean age 70.0, SD = 3.19) were recruited and were compensated $25 for their time. Data from two participants were removed because their performance on the math portion of the working memory test was less than 85%. Older adult characteristics are illustrated in Table 1.

[Table 1 about here]

2.1.2 Materials

Equipment. PC-compatible (Windows 7) computers running at 3.2 GHz with 4GB of RAM were used with a 19-inch LCD monitor set at a resolution of 1024x1280 pixels. Participants were seated approximately 18 inches from the monitor and interacted primarily with a mouse (on the preferred side) and a keyboard.

Targeting Task. The task was a low-fidelity targeting simulation (Figure 1) used in previous studies ([blinded], in review; Rovira, McGarry, & Parasuraman, 2007). The task screen consisted of four areas: an overhead terrain grid (right), a target input area with automated assistance (left), and a communications module (top-left). The terrain grid displayed red enemy units (E1 to E3), green friendly units (A1 to A3), yellow friendly battalion units (B1 to B3), and one orange headquarters unit (HQ). The task was to assign the closest friendly unit to engage in combat with the most dangerous enemy target. The criteria for most dangerous enemy unit was whether it was closest to any friendly unit but also closest to HQ. Thus, if two enemy units were equally distant from any other friendly units, the enemy unit closest to HQ was deemed most dangerous. These criteria were derived from consulting military subject matter experts.

[Figure 1a and Figure 1b about here]

To aid participants with this task, the automation (lower-left) provided either a lower degree of automation (termed information-analysis automation; Figure 1a) or a higher degree of automation (decision-selection; Figure 1b). Information-analysis automation displayed all pairs of friendly-enemy distances in an arbitrary order, alleviating the participant from manually calculating distances but still requiring scanning and comparison of the values to determine best pairing. It is considered a lower degree of automation because it is automation at a lower stage: information automation (without a direct pointer to a decision) and lower level (a complete set of all calculated distances). Decision-selection automation also calculated the distances but ordered them by pair distance and narrowed the displayed options to the top three, alleviating a major source of working memory demand. It is considered a relatively higher degree of automation because it represents automation at a higher stage: decision automation (facilitates working memory by providing decision selections) and higher level (narrows the calculated distances to a few). When the automation was not faulty, the top choice in the table was always the best choice. These two degrees of automation were chosen because they straddle the critical boundary between information and decision automation where the negative performance effects of automation failures are most apparent in the latter (Onnasch, Wickens, Li, & Manzey, 2013). This is because the greater support provided by higher degrees of automation necessarily means that the user is less engaged in the task, more “out of the loop” and thus is more affected by the consequences of low situation awareness when faced with an automation failure (Endsley & Kiris, 1995). Since the focus of this research was on age-related differences with automation as it crosses the critical boundary between information and decision automation, we did not include a manual or unaided condition.

During the task, participants could either use the assistance of the automation or make their own enemy-friendly unit engagement decisions, but were required to respond within 20s. Participants were able to verify the automation by reviewing the terrain view and manually computing distances themselves by counting grid boxes. After they made their selection, or if 20 seconds had elapsed, the trial ended and the terrain map was replaced with a new grid of enemy, friendly, and HQ units. The effects of automation on performance and complacency manifest themselves most strongly in multitask situations (Parasuraman & Manzey, 2010) thus a secondary, non-automated task was included. The task was to monitor the communications panel for a call for communications (call sign) that appeared every 6s. Participants were required to click on the ANSWER button every time their personal call sign appeared while they were selecting units. This secondary task was always performed; there was no single task condition.

Working memory measure. Working memory span was measured using the automated operation span task (AOSPAN), a computerised version of the operation span memory task (Unsworth, Heitz, Schrock, & Engle, 2005). In the AOSPAN task, participants are shown sequences of letters, one at a time. In between the letter presentation, they were given a math problem and had to determine whether it was correct or not (Figure 2). Participants were instructed to maintain at least an accuracy rate of 85% on the math task. After the presentation of a set of letters and math problems, participants had to recall the letters in the correct display sequence using a mouse. Running performance feedback on the letter recall and math verification task was provided after each response.

[Figure 2 about here]

Automation Attitude and Trust Measures. Pre-existing individual differences in attitudes toward automation were measured with the Complacency Potential Rating Scale (CPRS; Singh, Molloy, & Parasuraman, 1993). CPRS is a 16-item test designed to measure the potential for complacency to common examples of automation (e.g., automated teller machines). Participants responded to the extent they agreed with statements about automation on a scale of 1 to 5. The CPRS score was a sum of these responses and ranged from 16 (low complacency potential) to 80 (high complacency potential. Trust was measured using a 12 item questionnaire adapted from Jian, Bisantz, and Drury (2000). The first five questions assessed negative attitudes toward automation while the remaining 7 assessed positive attitudes. Trust was also measured after every block by asking four questions adapted from Lee and Moray (1994). The questions assessed the extent to which participants rated their trust in the automation, their self-confidence during the task, the degree to which they relied on the automation, and their belief that the automation improved their performance. Participants responded by using an on-screen visual analogue scale ranging from 0 to 100.

2.1.3 Design and Procedure

The experiment was a 2 (degree of automation: information (low DOA), decision (high DOA)) x 2 (automation correctness: correct, incorrect) within-subjects design. Participants completed eight blocks of 20 trials each for a total of 160 trials. The overall automation reliability was set at 80%; that is, in a block of 20 trials, the automation provided correct assistance on 16 trials while providing incorrect assistance on 4 trials. In an incorrect assistance trial, the automation provided faulty enemy-friendly distance information, thus the distance and best-pairing order (with decision automation) were incorrect. The first automation failure did not occur until the 10th trial, so that users would not discount the automation (Wickens & Xu, 2002). Subsequent automation failures were distributed randomly throughout the remaining trials. Presentation of each automation degree (information analysis or decision selection) was blocked so that each block only presented one degree of automation, and counterbalanced. History-based trust was assessed after every block. The dependent variable was decision accuracy of enemy-friendly engagement selections. Decision accuracy was calculated as the percentage of trials in which the participant correctly selected the optimal enemy-friendly pairing.

Participants were tested in groups of up to 6 at a time on individual computers with a mouse on their preferred side. After signing the informed consent, participants started the cognitive ability tests. After all participants completed the tests, they were given instructions about the decision-making task. They were also told that the automated technology provided was very reliable but it was not perfect all the time. No other information about the reliability was given. After these instructions, the participants started 8 trials of practice (automation was correct on all practice trials: 4 information-analysis and 4 decision-selection). Once the practice trials were completed, the experimenter answered any questions before the participants started the actual task. After completing the task, participants filled out the CPRS and trust questionnaire.

2.2 Results

To examine whether participants devoted a differential amount of effort to the secondary communications task between the automation conditions, communications task accuracy data was subjected to an analysis of variance (ANOVA). There was no significant difference in communications task accuracy by automation condition (information $M = 0.52$, $SD = 0.27$, decision $M = 0.51$, $SD = 0.28$, $F(1,167)=0.01$, $p > .05$).

2.2.1 Correlations between Decision Accuracy, Trust, and Working Memory

We first computed correlations between trust and CPRS measures. (Table 2). The correlation table shows that the history-based questions were correlated with each other in the expected direction (i.e., trust was positively correlated with extent to which participants “relied”). As expected, positive and negative trust were negatively correlated, $r = -.32$. Finally, CPRS was not related to any history-based trust question or positive and negative trust, indicating that it could be a construct unique from specific attitudes toward automation (positive/negative trust). Correlations between decision accuracy, trust, and working memory were computed (Table 3). Positive trust was negatively correlated with accuracy when the automation failed, $r = -.43$, $r = -.47$, respectively, indicating that more positive attitudes toward automation were associated with lower accuracy when the automation failed (an indicator of complacency).

[Table 2 about here]

[Table 3 about here]

Decision accuracy on automation-correct trials with information automation was not related to any trust measure. However, decision accuracy on automation-failure trials with information automation was negatively correlated to their responses to the history-based questions of 1) “how much did you trust”, 2) how much did you rely”, and 4) “how much did the automation improve performance”, r = -.62, r = -.58, r = -.58, respectively, indicating that high levels of trust were associated with worse performance when the automation failed, indicative of bias or complacency (over trust). Ratings of self confidence (question 3) were significantly positive correlated to accuracy when the automation failed, r = 0.55, indicating that higher self-confidence was associated with greater accuracy when the automation failed.

Decision accuracy on automation-correct trials with decision automation was significantly correlated with trust, reliance, and beliefs that automation helped performance, r = .37, r = .36, r = .37, respectively. When decision automation failed, trust, reliance, and the belief that automation helped all negatively correlated with accuracy, r = -.64, r = -.65, r = -.66, respectively, again, indicative of the detrimental effects of complacency. Self-confidence was positively correlated to performance when the automation failed, r = .56.

Decision accuracy when the automation was correct in information automation and decision automation was significantly correlated with working memory, r = .44 and r = .37, respectively, indicating that higher working memory was associated with greater accuracy. We next probed the relationship between individual differences in working memory and decision accuracy as a function of automation factors (degree and automation correctness).

2.2.2 Multilevel models to further examine role of working memory

Multilevel modelling (MLM) was employed to examine the influence of individual differences in working memory, degree ofautomation, and automation correctness on decision accuracy. MLM was preferred over regression for several reasons. First, MLMs are ideal for nested data structures (i.e., repeated measures), such as automation-correct and automation-failure trials within a participant, as MLMs simultaneously estimate both intra- and inter-individual (Raudenbush & Bryk, 2002). Second, as our overall automation reliability was 80% (128 trials where the automation was correct, 32 trials where the automation failed per participant) there was a large difference in the number of observations of this independent variable which is a problem with repeated measures regressions but not for MLM (Neupert et al., 2006a, 2006b). Third, regression is more likely to produce Type I errors with nested data, making MLM a more conservative choice (Hox & Bechger, 1998; Raudenbush & Bryk, 2002; Tabachnick & Fidell, 2007). Finally, MLM allows for the examination of critical cross-level interactions, such as the interaction between the person-level predictor of working memory ability and the trial-level predictors of automation degree and automation correctness. Multilevel modelling was implemented using PROC MIXED through SAS, version 9.4.

[Table 4 about here]

A two-level hierarchical model assessed the effects of the within-person variables (level 1) of degree of automation and automation correctness, the between-person predictor (level 2) of working memory score, and their cross-level interactions on decision accuracy in the task (Table 4). We used a model-building approach where we first ensured there was significant variability at both levels to allow predictors to be entered at those levels (Model 1), then added the within-participant predictors manipulated in the task (Model 2), and finally added the between-participant predictor of working memory score and controlled for CPRS score (Model 3). Each model was compared to the previous using Akaike's information criterion (AIC) values to ascertain if the added predictors increased the quality of the model. Lower AIC values indicate better model fit.

Model 1: No predictors. The first step was to run a fully unconditional model, one without any predictors (Table 4: Model 1), to discover how much variance in decision accuracy was due to within- ($\sigma 2$) and between-participant ($\tau 00$) differences and whether that variance was significantly high enough to try and explain using our predictors. The unconditional model revealed significant variance at both levels: 90% of the variance was at the within-person level ($\sigma 2 = 0.211$, $z = 54.94$, $p < .0001$) and 10% of the variance was at the between-person level ($\tau 00 = 0.023$, $z = 4.16$, $p < .0001$).

Model 2: Within-person variables. Model 2 revealed that automation correctness and degree of automation explained 7% of the within-subject variance. Model fit using the Akaike's information criterion (AIC) improved from 7903.8 to 7472.3. The model revealed significant main effects of automation degree, $F(1,6033) = 2.14$, $p = .033$, indicating that participant accuracy was greater with decision automation ($M = 0.59$, $SD = 0.30$) compared to information automation ($M = 0.50$, $SD = 0.25$). The main effect of automation correctness was also significant, $F(1,6033) = 12.41$, $p < .0001$), indicating that accuracy was higher on automation-correct trials ($M = 0.67$, $SD = 0.21$) versus automation-failure trials ($M = 0.42$, $SD = 0.29$). The two-way interaction of degree of automation and automation correctness was also significant. $F(1,6033) = 2.04$, $p = .042$, illustrated in Figure 3. Pairwise comparisons (Sidak-adjusted for multiple comparisons) to decompose the two-way interaction showed that when the automation failed, decision accuracy was not significantly different between information automation (M = 0.3, SD = 0.26) and decision automation (M = 0.44, SD = 0.30), but when the automation was correct, accuracy was significantly higher, $F(1,41) = 43.7$, $p < .05$, with decision automation (M = 0.74, SD = 0.21) compared to information automation (M = 0.61, SD = 0.20).

[Figure 3 about here]

These results partially support "conventional wisdom" (Onnasch, Wickens, Li, & Manzey, 2014), based on studies of younger adults, that higher degrees of correct automation increasingly benefit performance but higher degrees of automation, when the automation fails, increasingly harm performance. We observed the performance-enhancing effects of higher degrees of correct automation in older adults but not the greater performance decrement from higher degrees of automation when it failed. Model 3 was added to examine the contribution of working memory ability on performance as a function of degree and automation correctness.

Model 3: Cross-level interactions. We expected older adults with lower working memory ability to depend more on the automation, and thus perform more poorly on trials where the automation failed compared to older adults with higher working memory ability. A third model was conducted to include working memory ability to examine this hypothesised cross-level interaction.

Model 3 returned similar effects to Model 2 regarding the within-participant predictors. A cross-level interaction was also present, where working memory ability interacted with automation correctness, $F(1,6030) = 20.68$, $p < .0001$) but not degree of automation. Contrasts showed that, contrary to expectations, when the automation failed, low and high working memory participants' accuracy did not significantly differ, $t(6030)=0.71$, $p=.478$, but when the automation was correct, higher working memory participant's accuracy was higher than lower working memory participants, $t(6030)=3.83$, $p <.0001$. Both low and high working memory groups had significantly higher accuracy when the automation was correct compared to when it failed (low working memory participants: $t(6030)=9.00$, $p <.0001$; high working memory participants: $t(6030)=18.83$, $p <.0001$). Model fit using AIC improved from 7472.3 to 7471.8, indicating the benefit of including working memory ability as a predictor.

The results of Model 3 showed that working memory ability moderated performance on automation-correct trials. Older adults with higher working memory performed better when the automation was correct compared to those with lower working memory. But working memory was unrelated to accuracy when the automation failed. This is partially consistent with the results of [blinded] (in review) who found that with correct information automation, high working memory younger adults outperformed those with low working memory. We observed the same pattern but regardless of degree of automation. [blinded] found that when the automation failed, high working memory younger adults were able still maintain a performance advantage compared to low working memory younger adults. We did not observe this benefit of high working memory; all older adults, regardless of working memory level, performed poorly when the automation failed. This suggests that the reduced level of working memory ability in older adults (even at the high end) compared to younger adults was unable to compensate for automation failures.

3.0 Discussion

The present study was designed to address two significant gaps in the literature on aging and automation: 1) how does working memory affect performance with automation in older adults, and 2) how does working memory interact with degree of automation and automation failures to affect performance in older adults? Examining the ability/performance relationships showed that for older adults, working memory was significantly related to performance but in contrast to previous research, independent of degree of automation. In addition, higher span older adults' performance benefited from correct automation whereas lower span older adults did not. However, when the automation failed, all older adult's performance, regardless of working memory span, declined.

How do we interpret these findings in light of [blinded]'s (in review) experiment with younger adults? When the automation failed, [blinded] found that the higher span younger adults were able to maintain a performance advantage over the lower span group; that is, the low spans' performance declined with automation failures, but the high span group declined less. They interpreted this as a kind of "buffering effect" of having high working memory against the negative performance effects of failures of automation.

Working memory is only expected to exert an influence on performance only when the task requires controlled processing, or the ability to override automatic tendencies (Unsworth & Engle, 2007; Rosen & Engle, 1998). This suggests that the nature of the buffering effect found in [blinded]'s high working memory group might be an enhanced ability to deploy controlled processing resources or to suppress intrusive thoughts or behaviours (Conway, Cowan, & Bunting, 2001; Kane, Bleckley, Conway, & Engle, 2001; Cantor & Engle, 1993). Specifically, in the context of facing an automation failure, having higher working memory abilities could mean an enhanced ability to detect the automation failure, the ability to manually carry out the task without automation, or an increased ability to suppress an automatic bias to comply with automation (Mosier & Skitka, 1996). Since we did not observe a performance difference (on trials where the automation failed) between low and high working memory older adults, it suggests that the buffering effect of high working memory observed by [blinded] can be localised to a component of working memory, or other mediator, that is sensitive to age-related declines (e.g., processing speed versus storage capacity; Salthouse & Babcock, 1991). Future research should further isolate the exact component of working memory that is responsible for the buffering effect.

The current results also reinforce the importance of perceptions of automation (e.g., trust) in addition to cognitive ability factors in explaining age-related differences in performance (Czaja, et al., 2006). As the ability/performance correlations showed, trust, in addition to working memory, was significantly associated with performance in older adults. But can age-related differences in the perception of automation, and the task environment, also explain older adults' characteristic behaviour with automation? In an experiment examining reliance on human versus automation, Lyon and Stokes (2012) found that operators relied more on automation over human collaborators when the perceived risk of the task was higher compared to lower. Extrapolating their results to aging, can older adults' reliance on automation, and inability to compensate during failure, be explained by their risk perceptions when interacting with automation? Older adults have reduced familiarity with technology compared to younger adults (Olson, O'Brien, Rogers, & Charness, 2011), thus, they may perceive most interactions with automated technology as relatively higher risk or uncertain leading to their increased reliance on automation. Additionally, slowing of perception, and cognitive processes associated with aging (Salthouse, 1991) may cause older adults to perceive increased time-pressure in situations where younger adults would not. It is for this reason that allowing older adults increased time is one of the most suggested age-related design recommendations (e.g., Fisk, Rogers, Charness, Czaja, & Sharit, 2009; Kelley &

Charness, 1994; Mayhorn, Stronge, McLaughlin, & Rogers, 2004; Pak & McLaughlin, 2011).

The idea that users, including older adults, may rely more on pre-existing and automatic biases or heuristics (such as automation heuristics; Mosier & Skitka, 1996) under situations of uncertainty was found in an experiment by Pak, McLaughlin, and Bass (2014). Thus, age-related complacency and bias with automation may not be a direct consequence of low working memory, but may be mediated by older adult's over-use of, or inability to inhibit a "trust technology" heuristic in situations of uncertainty or time pressure; an adaptive strategy when resources are low or risk perception is high, to guide behaviour (Bahner, Hüper, & Manzey, 2008; Mosier & Skitka, 1996).

4.0 Conclusion

Although higher degrees of automation, when that automation is correct, are beneficial for low-span young adults (e.g., Onnasch, Wickens, Li, & Manzey, 2014), it is automation correctness, rather than degree, that seems most crucial for older adults. Short of simply increasing the reliability of automation that will be used by older adults, these results suggest that the design of automation intended for use by older adults should enhance the display of the likelihood the automation is correct but also assist the older user in recovering from automation failure. The importance of automation correctness is conceptually consistent with research that has shown that older adults are keenly aware of the reliability of technology and that it negatively affects their perceptions of technology (Mitzner et al., 2010). Prior research has shown that the display of such information is beneficial for trust calibration and reducing automation bias in younger adults (McGuirl & Sarter, 2006). The current results suggest that such a display might be especially helpful to older adults by alleviating the working memory demands of recovering from a failure of the automation (environmental support;

Morrow & Rogers, 2008). Promising research suggests that providing younger adults with reliability information that is artificially higher than normal might actually *properly* calibrate beliefs and expectations about automation (Barg-Walkow & Rogers, 2015). Such a strategy may also work with older adults because it may take advantage of their somewhat increased tendency to trust automation but takes into account their tendency to estimate reliability downward (Sanchez, Fisk, & Rogers, 2004). Their mis-calibration of reliability may come from the inability to remember and aggregate past automation performance due to working memory capacity limitations. One of the more unexpected findings was that we did not observe a larger performance decrement with decision automation compared to information automation. A meta-analysis of automation studies (Onnasch, Wickens, Li, & Manzey, 2014) provided support for the notion that higher degrees of automation, when that automation is correct, increasingly help performance but that higher degrees of automation, when that automation fails, should increasingly harm performance. The rationale is that with higher degrees of automation, where the task is allocated more to automation, the user is likely to suffer from out-of-the-loop unfamiliarity and is thus unprepared to resume manual control when automation fails. We found that older adults benefited from higher degrees of correct automation, consistent with the literature, but they were not increasingly harmed by higher degrees of automation, when that automation failed. The lack of greater performance harm from automation failures at higher degrees of automation is unlikely to be a floor effect as all older adults were still able to maintain a level of performance well above chance[1]. The detrimental effects of different degrees of automation

[1] Participants were presented with 160 total trials (80 information automation, 80 decision automation). Information automation presented 9 possible pairings (1/9 chance of choosing correct pairing) while decision automation presented 3 possible pairings (1/3

typically become most evident under higher workload (Parasuraman & Manzey, 2010), thus a future study might examine older adults' interaction with automation in more demanding situations to fully visualise the effect of higher degrees of automation on older adults' performance.

Given the finding that degree of automation is less impactful of older adults' performance with automation compared to its correctness, another recommendation for the design of automation intended for older adults might be to provide a higher degree of automation rather than a lower degree. The logic of this recommendation is that the older adults in our sample did not seem to be subject to the "lumberjack effect" (increased performance with automation of higher degree that is reliable, but catastrophic performance with automation of higher degree that is unreliable; Onnasch et al., 2014) as younger adults were. Older adults received the positive benefits of higher degrees of correct automation but curiously were not subject to the negative effects of incorrect, higher degree automation.

Another limitation of our experiment was that the automation paradigm used in this experiment was not able to differentiate whether a response reflected compliance and reliance behaviour. Compliance is a behaviour where the operator adheres to unreliable automation while reliance is where the operator does not act (verify) when the automation fails to alert. While the concepts are typically studied most often with alerting automation, the concepts have an equivalent in some forms of decision automation as well (commission versus omission errors; Manzey, Reichenbach, & Onnasch, 2012; Skitka, Mosier, & Burdick, 1999). Given that compliance and reliance

---

chance). But within each of those 80 trials, automation failed on 16 trials (64 correct; 80% reliability). Thus, chance performance in this study was calculated to be, [(1/9) * 0.8 + (1/3) * 0.8] / 2 = 0.18

behaviours may be theoretically unique (Meyer, 2004) and have different causes and effects (Dixon, Wickens, & McCarley, 2007), a future experiment might examine the role of age-related differences in working memory and other abilities specifically on older adult's pattern of compliance and reliance behaviours.

5.0 References

Ahmed, Nisar, Ewart de Visser, Tyler Shaw, Amira Mohamed-Ameen, Mark Campbell, and Raja Parasuraman. "Statistical modelling of networked human-automation performance using working memory capacity." Ergonomics 57, no. 3 (2014): 295-318.

Bahner, J. E., A. D. Hüper, and D. Manzey. 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.

Barg-Walkow, Laura H., and Wendy A. Rogers. "The effect of incorrect reliability information on expectations, perceptions, and use of automation." *Human Factors* (2015): 0018720815610271.

Cantor, J., and R. W. Engle. 1993. "Working-Memory Capacity as Long-Term Memory Activation: An Individual-Differences Approach." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19 (5): 1101-1114. doi:10.1037/0278-7393.19.5.1101.

Conway, A. R., N. Cowan, and M. F. Bunting. 2001. "The Cocktail Party Phenomenon Revisited: The Importance of Working Memory Capacity." *Psychonomic Bulletin & Review* 8 (2): 331-335. doi:10.3758/BF03196169.

Craik, F. I., and T. A. Salthouse. 2011. *The Handbook of Aging and Cognition*. New York, NY: Psychology Press.

Crocoll, W. M., and B. G. Coury. 1990. "Status or Recommendation: Selecting the Type of Information for Decision Aiding." In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 34 (19): 1524-1528. doi:10.1177/154193129003401922.

Czaja, S. J., N. Charness, A. D. Fisk, C. Hertzog, S. N. Nair, W. A. Rogers, and J. Sharit. 2006. “Factors Predicting the Use of Technology: Findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). *Psychology and Aging*, 21 (2): 333-352. doi:10.1037/0882-7974.21.2.333.

de Visser, E., T. Shaw, A. Mohamed-Ameen, and R. Parasuraman. 2010. “Modeling Human-Automation Team Performance in Networked Systems: Individual Differences in Working Memory Count.” In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 54 (14): 1087-1091. doi:10.1177/154193121005401408.

De Winter, J. C. F., D. Dodou, and P. A. Hancock. "On the paradoxical decrease of self-reported cognitive failures with age." *Ergonomics* 58, no. 9 (2015): 1471-1486.

Dixon, S. R., and C. D. Wickens. 2007. “On the Independence of Compliance and Reliance: Are Automation False Alarms Worse Than Misses?.” *Human Factors* 49 (4): 564-572. doi:10.1518/001872007X215656.

Donmez, B., L. N. Boyle, J. D. Lee, and D. V. McGehee. 2006. “Drivers’ Attitudes Toward Imperfect Distraction Mitigation Strategies.” *Transportation Research Part F: Traffic Psychology and Behaviour* 9 (6): 387-398. doi:10.1016/j.trf.2006.02.001.

Endsley, M. R., and E. O. Kiris. 1995. “The Out-of-the-Loop Performance Problem and Level of Control in Automation.” *Human Factors* 37 (2): 387–394. doi:10.1518/001872095779064555.

Engle, R. W., S. W. Tuholski, J. E. Laughlin, and A. R. Conway. 1999. “Working Memory, Short-Term Memory, and General Fluid Intelligence: A Latent-Variable Approach.” *Journal of Experimental Psychology: General* 128 (3): 309-331. doi:10.1037/0096-3445.128.3.309.

Ezer, N., A. D. Fisk, and W. A. Rogers. 2008. “Age-Related Differences in Reliance Behavior Attributable to Costs Within a Human-Decision Aid System.” *Human Factors* 50 (6): 853-863. doi:10.1518/001872008X375018.

Fisk, A. D., W. A. Rogers, N. Charness, S. J. Czaja, and J. Sharit. 2009. *Designing for Older Adults: Principles and Creative Human Factors Approaches*. Boca Raton, FL: CRC press.

Fox, J., and D. Boehm-Davis. 1998. “Effects of Age and Congestion Information Accuracy of Advanced Traveler Information Systems on User Trust and

Compliance." *Transportation Research Record: Journal of the Transportation Research Board* 1621: 43-49. doi:10.3141/1621-06.

Gilbert, D. K., and W. A. Rogers. 1999. "Age-Related Differences in the Acquisition, Utilization, and Extension of a Spatial Mental Model." *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 54 (4): 246-255. doi:10.1093/geronb/54B.4.P246.

Hancock, Peter A. "Automation: how much is too much?" *Ergonomics* 57, no. 3 (2014): 449-454.

Ho, G., L. M. Kiff, T. Plocher, and K. Z. Haigh. 2005. "A Model of Trust & Reliance of Automation Technology for Older Users." In *AAAI-2005 Fall Symposium: Caring Machines: AI in Eldercare*: 45-50.

Ho, G., D. Wheatley, and C. T. Scialfa. 2005. "Age Differences in Trust and Reliance of a Medication Management System." *Interacting with Computers* 17 (6): 690–710. doi:10.1016/j.intcom.2005.09.007.

Hoff, K. A., and M. Bashir. 2015. "Trust in Automation Integrating Empirical Evidence on Factors that Influence Trust." *Human Factors* 57 (3): 407-434. doi:10.1177/0018720814547570.

Hox, J. J., and T. M. Bechger. 1998. "An Introduction to Structural Equation Modeling." *Family Science Review* 11: 354-373.

Jian, J., A. M. Bisantz, and C. G. Drury. 2000. "Foundations for an Empirically Determined Scale of Trust in Automated Systems." *International Journal of Cognitive Ergonomics* 4 (1): 53-71. doi:10.1207/S15327566IJCE0401_04.

Kane, M. J., M. K. Bleckley, A. R. Conway, and R. W. Engle. 2001. "A Controlled-Attention View of Working-Memory Capacity." *Journal of Experimental Psychology: General* 130 (2): 169-183. doi:10.1037/0096-3445.130.2.169.

Kelley, C. L., and N. Charness. 1995. "Issues in Training Older Adults to Use Computers." *Behaviour & Information Technology* 14 (2): 107-120. doi:10.1080/01449299508914630.

Kieras, D. E., and S. Bovair. 1984. "The Role of a Mental Model in Learning to Operate a Device." *Cognitive Science* 8 (3): 255-273. doi:10.1207/s15516709cog0803_3.

Lee, J. D., and N. Moray. 1994. "Trust, Self-Confidence, and Operators' Adaptation to Automation." *International Journal of Human-Computer Studies* 40 (1): 153-184. doi:10.1006/ijhc.1994.1007.

Lee, J. D., S. Stone, B. F. Gore, C. Colton, J. Macauley, R. Kinghorn, J. L. Campbell, M. Finch, and G. Jamieson. 1997. *Advanced Traveler Information Systems and Commercial Vehicle Operations Components of the Intelligent Transportation Systems: Design Alternatives for In-Vehicle Information Displays*. No. FHWA-RD-96-147.

Lyons, J. B., and C. K. Stokes. 2012. “Human-Human Reliance in the Context of Automation.” *Human Factors* 54 (1): 112–121. doi:10.1177/0018720811427034.

Manzey, D., J. Reichenbach, and L. Onnasch. 2012. “Human Performance Consequences of Automated Decision Aids: The Impact of Degree of Automation and System Experience.” *Journal of Cognitive Engineering and Decision Making* 6 (1): 57-87. doi:10.1177/1555343411433844.

Mayhorn, C. B., A. J. Stronge, A. C. McLaughlin, and W. A. Rogers. 2004. “Older Adults, Computer Training, and the Systems Approach: A Formula for Success.” *Educational Gerontology* 30 (3): 185-203. doi:10.1080/03601270490272124.

McBride, S. E., W. A. Rogers, and A. D. Fisk. 2011. “Understanding the Effect of Workload on Automation Use for Younger and Older Adults.” *Human Factors* 53 (6): 672-686. doi:10.1177/0018720811421909.

McKendrick, Ryan, Tyler Shaw, Ewart de Visser, Haneen Saqer, Brian Kidwell, and Raja Parasuraman. "Team performance in networked supervisory control of unmanned air vehicles effects of automation, working memory, and communication content." Human Factors: The Journal of the Human Factors and Ergonomics Society (2013): 0018720813496269.

McGuirl, J. M. 2006. “Supporting Trust Calibration and the Effective Use of Decision Aids by Presenting Dynamic System Confidence Information.” *Human Factors* 48 (4): 656-665. doi:10.1518/001872006779166334.

Mitzner, T. L., J. B. Boron, C. B. Fausset, A. E. Adams, N. Charness, S. J. Czaja, K. Dijkstra, A. D. Fisk, W. A. Rogers, and J. Sharit. 2010. “Older Adults Talk Technology: Technology Usage and Attitudes.” *Computers in Human Behavior* 26 (6): 1710–1721. doi:10.1016/j.chb.2010.06.020.

Morrow, D. G., and W. A. Rogers. 2008. “Environmental Support: An Integrative Framework.” *Human Factors* 50 (4): 589–613. doi: 10.1518/001872008X312251.

Mosier, K. L., and L. J. Skitka. 1996. “Human Decision Makers and Automated Decision Aids: Made for Each Other.” *Automation and Human Performance: Theory and Applications*: 201-220.

Meyer, J. 2004. “Conceptual Issues in the Study of Dynamic Hazard Warnings.” *Human Factors* 46 (2): 196–204. doi:10.1518/hfes.46.2.196.37335.

Neupert, S. D., D. M. Almeida, D. K. Mroczek, and A. Spiro III. 2006a. “The Effects of the Columbia Shuttle Disaster on the Daily Lives of Older Adults: Findings from the VA Normative Aging Study.” *Aging and Mental Health* 10 (3): 272-281. doi:10.1080/13607860500409682.

Neupert, S. D., D. M. Almeida, D. K. Mroczek, and A. Spiro III. 2006b. “Daily Stressors and Memory Failures in a Naturalistic Setting: Findings from the VA Normative Aging Study.” *Psychology and Aging* 21 (2): 424-429. doi:10.1037/0882-7974.21.2.424.

Olson, K. E., A. D. Fisk, and W. A. Rogers. 2009. “Collaborative Automated Systems: Older Adults' Mental Model Acquisition and Trust in Automation.” In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 53 (22): 1704-1708. doi:10.1177/154193120905302210.

Olson, K. E., M. A. O’Brien, W. A. Rogers, and N. Charness. 2011. “Diffusion of Technology: Frequency of Use for Younger and Older Adults.” *Ageing International* 36 (1): 123–145. doi:10.1007/s12126-010-9077-9.

Onnasch, L., S. Ruff, and D. Manzey. 2014. “Operators′ Adaptation to Imperfect Automation–Impact of Miss-Prone Alarm Systems on Attention Allocation and Performance.” *International Journal of Human-Computer Studies* 72 (10): 772-782. doi:10.1016/j.ijhcs.2014.05.001.

Onnasch, L., C. D. Wickens, H. Li, and D. Manzey. 2014. “Human Performance Consequences of Stages and Levels of Automation: An Integrated Meta-Analysis.” *Human Factors* 56 (3): 476-488. doi:10.1177/0018720813501549.

Pak, R., and A. McLaughlin. 2010. *Designing Displays for Older Adults*. Boca Raton, FL: CRC Press.

Pak, R., N. Fink, M. Price, B. Bass, and L. Sturre. 2012. “Decision Support Aids with Anthropomorphic Characteristics Influence Trust and Performance in Younger and Older Adults.” *Ergonomics* 55 (9): 1059-1072. doi:10.1080/00140139.2012.691554.

Pak, R., A. C. McLaughlin, and B. Bass. 2014. “A Multi-Level Analysis of the Effects of Age and Gender Stereotypes on Trust in Anthropomorphic Technology by Younger and Older Adults.” *Ergonomics* 57 (9): 1277–1289. doi:10.1080/00140139.2014.928750.

Sanchez, Julian, Arthur D. Fisk, and Wendy A. Rogers. "Reliability and age-related effects on trust and reliance of a decision support aid." In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 48, no. 3, pp. 586-589. SAGE Publications, 2004.

Parasuraman, R., E. de Visser, M. K. Lin, and P. M. Greenwood. 2012. “Dopamine Beta Hydroxylase Genotype Identifies Individuals Less Susceptible to Bias in Computer-Assisted Decision Making.” *PLoS ONE* 7 (6): e39675. doi:10.1371/journal.pone.0039675.

Parasuraman, R., T. B. Sheridan, and C. D. Wickens. 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. doi:10.1109/3468.844354.

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

Park, D., and N. Schwarz. 2000. *Cognitive Aging: A Primer*. New York, NY: Psychology Press.

Plude, D. J., and J. A. Doussard-Roosevelt. 1989. “Aging, Selective Attention, and Feature Integration.” *Psychology and Aging* 4 (1): 98-105. doi:10.1037/0882-7974.4.1.98.

Raudenbush, S. W., and A. S. Bryk. 2002. *Hierarchical Linear Models* (2nd Ed.). Thousand Oaks, CA: Sage Publications.

Rosen, V. M., and R. W. Engle. 1998. “Working Memory Capacity and Suppression.” *Journal of Memory and Language* 39 (3): 418-436. doi:10.1006/jmla.1998.2590.

Rovira, E., A. Cross, E. Leitch, and C. Bonaceto. 2014. “Displaying Contextual Information Reduces the Costs of Imperfect Decision Automation in Rapid Retasking of ISR Assets.” *Human Factors* 56 (6): 1036–1049. doi:10.1177/0018720813519675.

Rovira, E., K. McGarry, and R. Parasuraman. 2007. “Effects of Imperfect Automation on Decision Making in a Simulated Command and Control Task.” *Human Factors* 49 (1): 76-87. doi:10.1518/001872007779598082.

Salthouse, T. A. 1992. *Mechanisms of Age-Cognition Relations in Adulthood*. Hillsdale, NJ: Lawrence Erlbaum.

Salthouse, T. A. 1996. “The Processing Speed Theory of Adult Age Differences in Cognition.” *Psychological Review* 103 (3): 403-428. doi:10.1037/0033-295X.103.3.403.

Salthouse, T. A., and R. L. Babcock. 1991. “Decomposing Adult Age Differences in Working Memory.” *Developmental Psychology* 27 (5): 763-776. doi:10.1037/0012-1649.27.5.763.

Sanchez, J., W. A. Rogers, A. D. Fisk, and E. Rovira. 2014. “Understanding Reliance on Automation: Effects of Error Type, Error Distribution, Age and Experience.” *Theoretical Issues in Ergonomics Science* 15 (2): 134-160. doi:10.1080/1463922X.2011.611269.

Saqer, H., E. de Visser, A. Emfield, T. Shaw, and R. Parasuraman. 2011. “Adaptive Automation to Improve Human Performance in Supervision of Multiple Uninhabited Aerial Vehicles Individual Markers of Performance.” In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 55 (1): 890-893. doi:10.1177/1071181311551185.

Sarter, N. B., R. J. Mumaw, and C. D. Wickens. 2007. “Pilots' Monitoring Strategies and Performance on Automated Flight Decks: An Empirical Study Combining Behavioral and Eye-Tracking Data.” *Human Factors* 49 (3): 347-357. doi:10.1518/001872007X196685.

Sheridan, T. B. 1992. *Telerobotics, Automation, and Human Supervisory Control*. Cambridge, MA: MIT Press.

Sheridan, T. B., and W. L. Verplank. 1978. “Human and Computer Control of Undersea Teleoperators.” Cambridge, MA: Man-Machine Systems Laboratory, Department of Mechanical Engineering.

Shipley, W. C. 1986. *Shipley Institute of Living Scale.* Los Angeles, CA: Western Psychological Services.

Skitka, L. J., K. L. Mosier, and M. Burdick. 1999. “Does Automation Bias Decision-Making?.” *International Journal of Human-Computer Studies* 51 (5): 991-1006. doi:10.1006/ijhc.1999.0252.

Singh, I. L., R. Molloy, and R. Parasuraman. 1993. “Automation-Induced “Complacency”: Development of a Complacency-Potential Scale.” *International Journal of Aviation Psychology* 3 (2): 111-122. doi:10.1207/s15327108ijap0302_2.

Tabachnick, B. G., and L. S. Fidell. 2007. “Multilevel Linear Modeling.” *Using Multivariate Statistics*: 781-857.

Unsworth, N., and R. W. Engle. 2007. “The Nature of Individual Differences in Working Memory Capacity: Active Maintenance in Primary Memory and Controlled Search from Secondary Memory.” *Psychological Review* 114 (1): 104-132. doi:10.1037/0033-295X.114.1.104.

Unsworth, N., R. P. Heitz, J. C. Schrock, and R. W. Engle. 2005. “An Automated Version of the Operation Span Task.” *Behavior Research Methods* 37 (3): 498-505. doi:10.3758/BF03192720.

Verhaeghen, P., and T. A. Salthouse. 1997. “Meta-Analyses of Age–Cognition Relations in Adulthood: Estimates of Linear and Nonlinear Age Effects and Structural Models.” *Psychological Bulletin* 122 (3): 231-249. doi:10.1037/0033-2909.122.3.231.

Wechsler, D. 1997. *Wechsler Memory Scale III*. (3rd Ed.). San Antonio, TX: The Psychological Corporation.

Wickens, C. D., H. Li, A. Santamaria, A. Sebok, and N. B. Sarter. 2010. “Stages and Levels of Automation: An Integrated Meta-analysis.” In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 54 (4): 389-393. doi:10.1177/154193121005400425.

Wickens, C. D., and X. Xu. 2002. “Automation Trust, Reliability and Attention.” (Tech. Rep. AHFD-02-14/MAAD-02-2). Savoy, IL: Aviation Research Lab.

Wiegmann, D., J. S. McCarley, A. F. Kramer, and C. D. Wickens. 2006. “Age and Automation Interact to Influence Performance of a Simulated Luggage Screening Task.” *Aviation, Space, and Environmental Medicine* 77 (8): 825-831.

Figure 1a. Information analysis automation. All friendly (A) and enemy (E) distances are calculated and presented but ordered arbitrarily by enemy label.

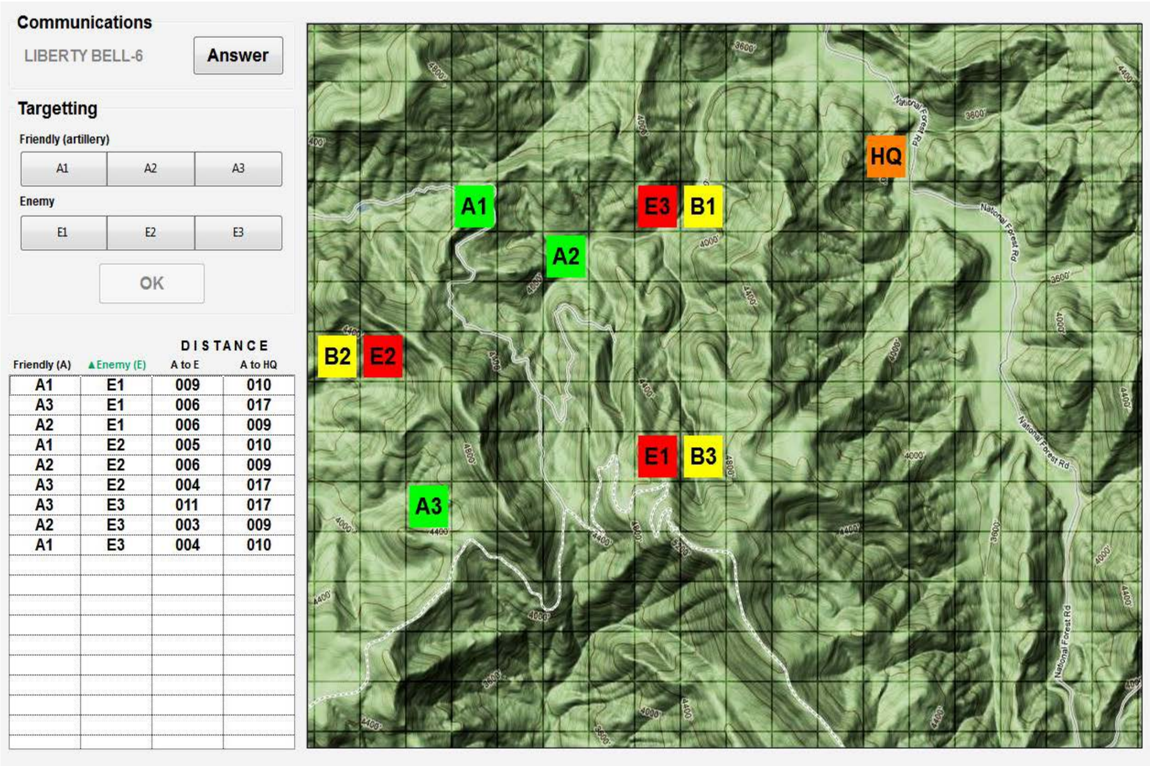

Figure 1b. Decision selection automation. All A to E distances are calculated but are ordered by distance and filtered to only show the top three.

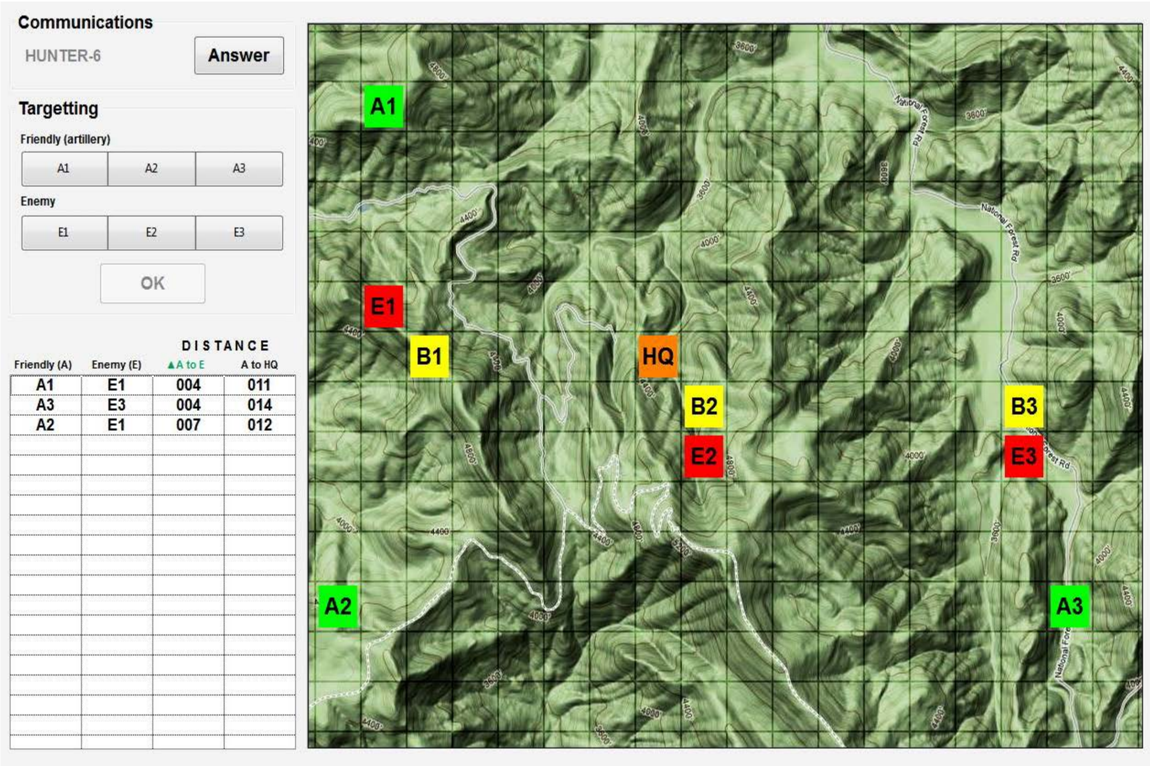

Figure 2. Operation span task sequence

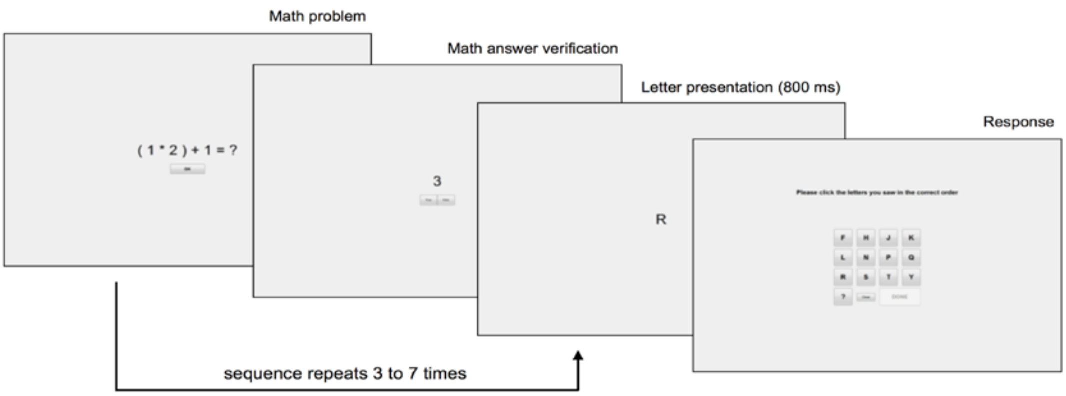

Figure 3. Accuracy as a function of degree of automation and correctness

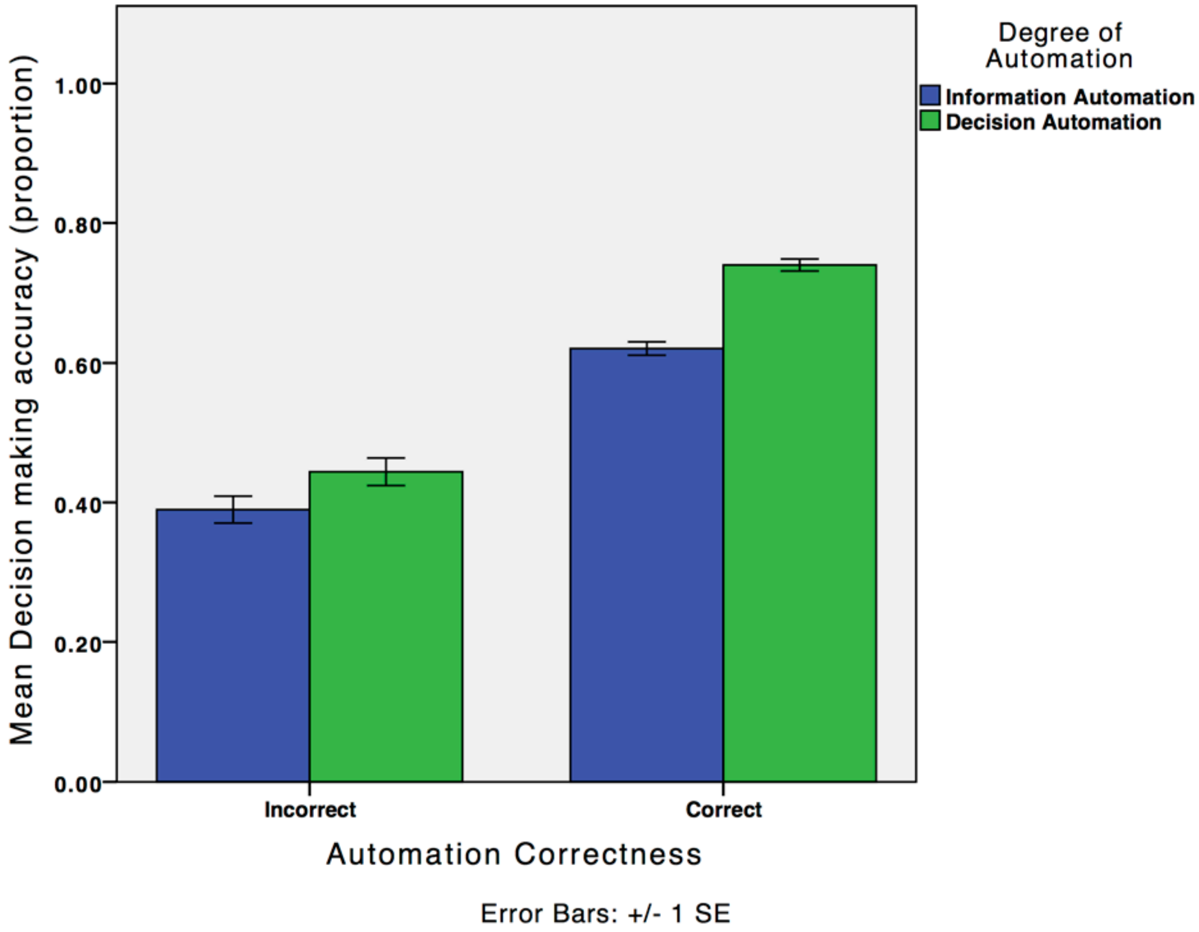

Table 1. Participant characteristics

	Mean	SD
Age	70	3.19
Working memory[a]	23.29	14.94
Shipley vocabulary[b]	36.02	2.31
Digit symbol substitution[c]	1773.19	422.1
Complacency potential[d]	46.83	3.66
Self reported health[e]	3.74	0.86

Note: [a]Total number of correctly recalled letter sets (maximum score 75; Unsworth, Heitz, Schrock, & Engle, 2005; higher is better). [b]Total correct (maximum score = 40; Shipley, 1986; higher is better). [c]Time to identify an incorrect trial in milliseconds (Wechsler, 1997; lower is better). [d]Tendency for complacency (Singh, Molloy, & Parasuraman, 1993; higher is more). [e]1 = Poor, 2 = Fair, 3 = Good, 4 = Very Good, 5 = Excellent

Table 2. Intercorrelations between trust and performance

		Information Analysis													
		Correct							Incorrect						
		1	2	3	4	5	6	7	1	2	3	4	5	6	7
1	How much did you trust	--							--						
2	How much did you rely	**0.75**	--						**0.75**	--					
3	Your self confidence	-0.29	**-0.58**	--					-0.29	**-0.58**	--				
4	How much did auto. improve perf.	**0.77**	**0.92**	**-0.53**	--				**0.78**	**0.92**	**-0.53**				
5	Complacency potential (CPRS)[b]	0.02	-0.12	-0.03	0.04	--			0.01	-0.12	-0.02	0.04	--		
6	Negative trust[c]	-0.05	0.08	-0.07	0.01	-0.25	--		-0.05	0.07	-0.07	0.00	-0.25	--	
7	Positive trust[c]	**0.63**	**0.44**	-0.18	**0.48**	0.13	**-0.32**		**0.63**	**0.45**	-0.19	**0.48**	0.13	**-0.32**	
8	Accuracy	0.25	0.24	-0.12	0.30	0.11	-0.15	0.16	**-0.62**	**-0.58**	**0.55**	**-0.58**	0.00	-0.08	**-0.43**
		Decision Selection													
		Correct							Incorrect						
1	How much did you trust	--							--						
2	How much did you rely	**0.88**	--						**0.89**	--					
3	Your self confidence	**-0.43**	**-0.47**	--					**-0.43**	**-0.48**	--				
4	How much did auto. improve perf.	**0.91**	**0.95**	**-0.52**					**0.91**	**0.96**	**-0.53**				
5	Complacency potential (CPRS)[b]	-0.02	-0.05	0.03	-0.03	--			-0.02	-0.06	0.04	-0.03	--		
6	Negative trust[c]	-0.01	0.04	-0.12	-0.06	-0.25	--		-0.01	0.04	-0.12	-0.05	-0.25	--	
7	Positive trust[c]	**0.55**	**0.45**	-0.14	**0.47**	0.13	**-0.32**		**0.54**	**0.44**	-0.14	**0.46**	0.13	**-0.32**	
8	Accuracy	**0.37**	**0.36**	-0.19	**0.37**	-.16	-0.21	0.29	**-0.64**	**-0.66**	**0.55**	**-0.66**	0.04	-0.15	**-0.47**

Note. Bolded items are significant, $p < .05$. [a]adapted from Lee & Moray (1994). [b]Tendency for complacency (Singh, Molloy, & Parasuraman, 1993; higher is more). [c]adapted from Jian, Bisantz, and Drury (2000).

Table 3. Intercorrelations and descriptive statistics of decision accuracy, trust, and working memory

		Information Analysis		Decision Selection					
		Correct	Incorrect	Correct	Incorrect	5	6	7	8
1	How much did you trust	0.25	**-0.62**	**0.37**	**-0.64**				
2	How much did you rely	0.24	**-0.58**	**0.36**	**-0.66**				
3	Your self confidence	-0.11	**0.55**	-0.19	**0.55**				
4	How much did auto. improve perf.	0.29	**-0.58**	**0.37**	**-0.66**				
5	Complacency potential (CPRS)[b]	0.11	0.00	0.16	0.04				
6	Negative dispositional trust[c]	-0.15	-0.08	-0.21	-0.15	-0.25			
7	Positive dispositional trust[c]	0.16	**-0.43**	0.29	**-0.47**	0.13	**-0.32**		
8	Working memory (OSPAN)[d]	**0.44**	-0.11	**0.37**	-0.11	0.22	-0.18	0.10	--
	Mean	0.62	0.39	0.73	0.45	46.80	14.90	26.70	23.29
	SD	0.20	0.26	0.21	0.30	3.70	7.30	9.20	3.19

Note. Bolded items are significant, *p* < .05. [a]adapted from Lee & Moray (1994). [b]Tendency for complacency (Singh, Molloy, & Parasuraman, 1993; higher is more). [c]adapted from Jian, Bisantz, and Drury (2000). [d]Total number of correctly recalled letter sets (maximum score 75; Unsworth, Heitz, Schrock, & Engle, 2005; higher is better).

Table 3. Unstandardised Coefficients of Multilevel Models of the Within- and Between-person Effects of Predictors on Decision Accuracy

	Model 1			Model 2			Model 3		
Fixed effects	Unconditional Model			Within-participant Predictors			Within- and Between-participant Predictors		
	Estimate		SE	Estimate		SE	Estimate		SE
Intercept	0.629	***	0.024	0.379	***	0.030	0.380	***	0.029
Between-person									
CPRS Score							0.002		0.007
Working Memory Score (WM)							-0.001		0.002
Within-person									
Automation Degree (AD)				0.054	*	0.025	0.054	*	0.025
Automation Correctness (AC)				0.250	***	0.020	0.249	***	0.020
AD x AC				0.058	*	0.028	0.058	*	0.028
Cross-level									
AD x WM							-0.001		0.002
AC x WM							0.006	***	0.001
AC x AD x WM							0.001		0.002
Random Effects									
$\sigma 2$	0.211		0.004	0.196		0.004	0.194		0.003
τoo	0.023		0.005	0.022		0.005	0.020		0.005
Model fit statistic									
AIC	7903.8			7472.3			7471.8		

Note. * $p < .05$, ** $p < .01$, *** $p < .001$; Working memory and CPRS scores were grand-mean centred. SE indicates standard error.