

Clemson University

**TigerPrints**

---

All Theses

Theses

---

May 2020

## The Effects of Increasing Degree of Unreliable Automation on Older Adults' Performance

Claire Textor

*Clemson University*, [ctextor@clemson.edu](mailto:ctextor@clemson.edu)

Follow this and additional works at: [https://tigerprints.clemson.edu/all\\_theses](https://tigerprints.clemson.edu/all_theses)

---

### Recommended Citation

Textor, Claire, "The Effects of Increasing Degree of Unreliable Automation on Older Adults' Performance" (2020). *All Theses*. 3349.

[https://tigerprints.clemson.edu/all\\_theses/3349](https://tigerprints.clemson.edu/all_theses/3349)

This Thesis is brought to you for free and open access by the Theses at TigerPrints. It has been accepted for inclusion in All Theses by an authorized administrator of TigerPrints. For more information, please contact [kokeefe@clemson.edu](mailto:kokeefe@clemson.edu).

THE EFFECTS OF INCREASING DEGREE OF UNRELIABLE AUTOMATION  
ON OLDER ADULTS' PERFORMANCE

---

A Thesis  
Presented to  
the Graduate School of  
Clemson University

---

In Partial Fulfillment  
of the Requirements for the Degree  
Master of Science  
Applied Psychology

---

by  
Claire Noel Textor  
May 2020

---

Accepted by:  
Dr. Richard Pak, Committee Chair  
Dr. Kelly Caine  
Dr. Ericka Rovira

## ABSTRACT

High level automation has the ability to relieve operators from complex, working memory-intensive tasks. When the task is primarily perceptual or cognitive in nature, the amount taken over by the machine can be very high. However, as operators interact with technology that is more automated (i.e., automation is higher in stage and degree), they may become more subject to the negative effects when that technology fails. This concept of reaping greater benefits of higher degrees of automation that is reliable but suffering catastrophic performance consequences when it is unreliable has been termed the lumberjack effect and has been well documented among younger adults (Endsley & Kiris, 1995; Onnasch et al., 2013; Rovira et al., 2017). The cause of this effect is that frequent interaction with reliable, high level automation induces a complacency or disengagement with the task (becoming out of the loop). Thus, when that automation fails, the user has been out of the loop (Endsley & Kiris, 1995) and is thus unprepared to resume the task. As older adults have reduced cognitive abilities, they may be even more subject to the lumberjack effect: benefiting greatly with reliable, high level automation but suffering major performance decrements with unreliable automation. The purpose of the current study was to examine the presence and magnitude of the lumberjack effect in older adults as it has not yet been documented in the literature. Older and younger adults interacted with various levels of automation. We replicated the finding that performance was negatively affected on unreliable trials of automation compared to reliable trials for both age groups (i.e., the lumberjack effect). However, this effect only appeared during low workload conditions and did not appear to be more pronounced in older adults. These results are the first to show that the lumberjack effect, previously observed in younger

adults is equally pronounced in older adults. However, what aspect of aging cognition was the source of this similar lumberjack effect is still an empirical question. Future work should be done to understand methods which can help older adults stay in the loop when using automated technology.

## ACKNOWLEDGEMENTS

I would first like to thank my advisor, Rich, for his unwavering patience and dedication. I would not be the researcher I am today without his devotion and mentorship. I would also like to thank my committee members, Dr. Kelly Caine and Dr. Ericka Rovira, for their guidance and positivity throughout this process. I can't put into words how grateful I am for my lab mates, Dan, Jeremy, Zach, and Heather. The conversations we share will last a lifetime. I would not be where I am today without the unwavering support of my parents and brother, thank you for always being there for me. Thank you to my amazing friends, for making me laugh when I want to cry and for putting up with my unending rants. And finally, thank you to my loving boyfriend Ruslan, for showing me nothing but care and compassion throughout this beautiful journey.

## TABLE OF CONTENTS

	Page
Title Page .....	i
ABSTRACT .....	ii
Acknowledgements .....	iv
List of Figures .....	vii
List of Tables .....	vii
INTRODUCTION .....	1
Types and Levels of Automation .....	2
Automation and the Lumberjack Effect .....	4
Older Adults and Automated Technology .....	5
Current Study .....	6
METHOD .....	8
Participants .....	8
Materials .....	9
Measures .....	13
Procedure .....	15
RESULTS .....	16
Subjective Workload .....	17

Trust and Automation.....	18
Working Memory Differences and Age.....	19
Task Time Between Age Groups .....	20
Secondary Task Performance .....	21
Taxi Dispatching Task and Decision Accuracy .....	21
<b>DISCUSSION.....</b>	<b>24</b>
Limitations and Future Directions.....	28
Conclusion.....	29
<b>APPENDICES .....</b>	<b>31</b>
Appendix A .....	32
Appendix B.....	33
Appendix C.....	34
Appendix D .....	35
Appendix E.....	36
Appendix F .....	37
Appendix G .....	39
Appendix H .....	40
Appendix I.....	43
Appendix J.....	45
<b>REFERENCES .....</b>	<b>46</b>

## List of Figures

Figure 1. Information analysis automation.....	11
Figure 2. Medium-decision automation .....	12
Figure 3. High-decision automation .....	13
Figure 4. Trust as a function of workload degrees of automation .....	19
Figure 5. Mean decision accuracy as a function of degree of automation and reliability for low workload conditions .....	23
Figure 6. Mean decision accuracy as a function of degree of automation and reliability for high workload conditions .....	24
Figure 7. History-based trust questionnaire.....	32
Figure 8. NASA-TLX.....	33
Figure 9. Mean ratings of subjective workload .....	35
Figure 10. Mean working memory differences between younger and older adults .....	36
Figure 11. Response time as a function of degree of automation and workload for older adults .....	37
Figure 12. Response time as a function of degree of automation and workload for younger adults....	38

## List of Tables

Table 1 Participant characteristics by age group including means and standard deviations for age in years, complacency potential rating scale (CPRS) ratings, and automation induced complacency potential (AICP) ratings. ....	17
Table 2 Mean scores for AICP, CPRS, and NASA-TLX for younger and older adults.....	34



Table 3 Independent sample t-tests between age groups comparing response times across all conditions broken down by automation reliability (unreliable or reliable), degree of automation (IA: information analysis, MD: medium decision, or HD: high decision) and workload (low or high)..... 41

Table 4 Independent sample t-tests between age groups comparing secondary task performance broken down by automation reliability (unreliable or reliable), degree of automation (IA: information analysis, MD: medium decision, or HD: high decision) and workload (low or high)..... 45

## INTRODUCTION

Automation is defined as a machine carrying out a function that was previously performed by a human (Parasuraman & Riley, 1997). Automation may help users by freeing them from mundane, repetitive, complex, or cognitively-demanding tasks. However, when automation is unreliable, the user is suddenly put in the situation of resuming manual control of a task for which they were not paying attention. A recent example of a high-level automation failure with dire consequences was the malfunctioning autopilot in the Boeing 737 MAX (“Pilots Warned”, 2019). The 737 MAX utilized a new auto pilot system that had not been fully tested. The plane erroneously descended at take-off, causing an alarm in the ground proximity warning system, another automated system, to alert. The pilot was not able to override the auto pilot in time and resulted in the death of all 157 people on board.

Higher level automation, as discussed above, tends to alleviate the operator from more working memory-intensive tasks such as decision making, compared to more perceptual tasks. However, as the Boeing story illustrates, there is a tradeoff: when the automation functions reliably, the operator has reduced working memory demand. However, when it malfunctions, the operator is left to diagnose a problem in a situation that they have not been attending (“out of the loop” phenomenon; OOTL) (Endsley & Kiris, 1995). When previously reliable automation fails, and the operator is put back “into the loop” of a complex, working memory-intensive task, performance typically suffers. This is part of what is known as the lumberjack effect (Onnasch et al., 2014). The lumberjack effect is the notion that when lower-levels of automation fail, the

consequences are less severe because the operator is still relatively engaged in the task (they were more “in the loop”). However, with higher level automation, they are more OOTL and thus, much less able to recover when automation fails.

### **Types and Levels of Automation**

While automation can take many forms (e.g., robots, software) it can be conceptually described by *type* (what task is being automated) and *level* (how much of a task is being automation. Automation can be categorized into four specific types, organized based on stages of human information processing that each supports: (1) information acquisition; (2) information analysis; (3) decision and action selection; (4) action implementation (Parasuraman, Sheridan, & Wickens, 2000). Each type exists on a continuum from levels of low to high automation. The combination of type and level is referred to as the degree of automation (Onnasch, 2014). The degree increases along with the level and type automation as a linear relationship. To understand degree of automation, it is necessary to distinguish between the four types.

The lowest and simplest type is information acquisition automation in which minimal processing is carried out on raw sensor-based data (e.g., camera). This data is then presented to the user, one example being a car’s backup camera. The driver is shown a live feed of what is going on behind the car while in reverse, but there is very little additional information. This type of automation serves to enhance the lower-level perceptual aspects of the task being performed.

In the next more complex type, information analysis automation, raw sensor data is processed or analyzed and presented to the user. An example of this type of automation might be a hypothetical night vision system in a car that not only enhances the

information (lightens the dark image) but also overlays an artificial path to indicate the hidden road and highlights important elements of the visual field (e.g., a deer in the road). Because the additional processing may alleviate the requirement for attentional selection, the user is naturally more OOTL compared to information acquisition automation.

The third type of automation, decision automation, processes and analyzes sensor data to a greater extent and presents choices that the user can compare. Decision automation, because of the additional analysis, alleviates a majority of the decision making components required to complete the task. While this type of automation does involve some processing of information, it only presents decision options to the user. A common example of this would be Google Maps which integrates multiple factors such as distance, means of transportation, and traffic conditions to calculate the fastest routes. It provides multiple suggestions and indicates how much faster one route would be compared to another, supplying additional details such as road closures and tolls, which allows the operator to make an informed decision. The operator is still able to make the final decision over which route to select and Google Maps will adjust its directions based on where the operator chooses to go, whether they follow a given route or deviate from it.

Action automation is the final type where the automation fully executes a task in place of the operator. A high level of this would be the autopilot in an airplane. A series of sensors detects information which is processed by the autopilot computer and once engaged, the pilot no longer has any role in the task as the plane maneuvers. The operator has the option of disengaging the autopilot, however, at such a high degree of automation, manual control would most likely be difficult if not impossible.

As the degree of automation increases, more of the task is allocated to the automation and less to the operator. Additionally, the nature of the task also changes as type of automation increases. With higher levels of automation, more of the cognitive components of the task are alleviated. This doesn't necessarily reduce workload, rather, it allows the operator to focus attentional resources elsewhere. This can be highly beneficial in domains which require attention to be allocated to multiple stimuli simultaneously.

### **Automation and the Lumberjack Effect**

Despite the benefits of reliable high-level automation, there are also potential negative effects (Bainbridge, 1983) such as the out-of-the-loop phenomenon. Operators who are out-of-the-loop for an extended period may experience skill degradation, loss of situation awareness, and increased complacency (Parasuraman, Sheridan & Wickens, 2000; Wickens et al., 1992). These consequences are only detrimental to performance when the automation fails and the operator must resume the task. Importantly, the level of the automation (i.e., amount of the task being assumed by automation) determines the consequences of failures such that failures of low-level automation may be of low magnitude while failures of high-level require cognitive effort for users to get back "in-the-loop." For example, malfunctions of a voice dictation system (a low-level system), which automates the well-learned skill of typing, merely requires users to resume typing, whereas failures of lane-keeping automation in a car (a high-level system) require operators to recognize the vehicles speed, location in the roadway, and proximity to potential hazards prior in order to safely take-over manual control of the vehicle.

The paradoxical notion that reliable high-level automation is extremely helpful to human performance but unreliable high-level automation is extremely detrimental to

human performance defines the lumberjack effect. Essentially, high-level automation can substantially reduce the amount of working memory demand that would otherwise be placed on the operator, allowing for cognitive resources to be freed and allocated to alternative tasks or stimuli. Unfortunately, when automation failure occurs, the operator is forced to quickly resume the task they were not attending to, thus regaining the working memory demand that was previously alleviated. This rapid switch from automated operation to manual control can lead to a sharp decline in task performance. Given that the sources of the lumberjack effect seem to be a) the unique cognitive properties of higher-levels of automation, which support more cognitive tasks, and b) the general difficulty of working-memory-intensive tasks, it is reasonable that individuals with diminished working memory capacity, such as older adults, may exhibit a stronger lumberjack effect.

### **Older Adults and Automated Technology**

Automation can be particularly beneficial to older adults who may experience age-related declines in many physical and cognitive abilities (e.g., Salthouse, 1994; Salthouse, 1996; Dobbs & Rule, 1990; Rybash et al., 1995). Given that the source of the lumberjack effect seems to be tied to working memory demands, studies using varying degrees of automation should show a relatively large lumberjack effect in older adults. However, this has not been the case. Pak et al. (2016) found that when automation was reliable, as expected, older adults' performance was enhanced. However, they did not observe a performance decrease when automation failed, as was expected. That is, they did not seem to observe a lumberjack effect in older adults.

There are some possible reasons that this study did not show a lumberjack effect, despite it being a well-supported phenomenon. First, the authors held workload constant at a low level so participants might have been able to easily recover from failure without the use of automation. Second, Pak et al. (2016) only used two levels of automation (information analysis, medium-decision). Recall that high level automation alleviates working memory demands more than lower levels. By using a moderate level of automation, participants, young and old, may have been able to easily recover from automation failure because manual calculation was relatively uncomplicated. Implementing a higher level of decision automation may increase OOTL effects, thus illustrating an increased lumberjack effect. Third, this task was carried out in a military domain which may not be familiar to many civilians. Performance could have been affected by the relative novelty of the domain if it was distracting or disorienting. If participants were not comfortable interacting with the system, they might have been more inclined to rely on themselves as opposed to the automation.

### **Current Study**

For the present study, we address these possible explanations to examine older adults use of automation. First, the most significant aspect of our study compared to Pak et al. (2016) is that we include a condition with an even higher level of automation. We expect that this higher degree will take the participant even more out of the loop, thus increasing the lumberjack effect. Second, in contrast to other studies (Rovira et al., 2017; Rovira, McGarry, & Parasuraman, 2007; Pak et al., 2016), we are utilizing a more conventional, civilian task domain. We predict that civilian participants might not have been accustomed to looking at terrain or had previous knowledge of UAVs. This could

have led to confusion and a lower likelihood of utilizing the automation aid. We anticipate that the use of taxis instead of UAVs will clear up any distraction or disorientation that may have resulted from the previous task domain. Third, we are manipulating workload, unlike Pak et al., (2016) who kept workload consistent throughout. We suspect that low task load (not enough working memory demand on the participant) may have been the reason they did not observe a lumberjack effect. Manipulating workload and increasing the degree of automation should allow us to more precisely control the working memory demand of the task.

The purpose of the current study was to examine how the lumberjack effect manifests in older adults. Recent findings have shown that lower working memory capacity is related to lower task performance (Rovira et al., 2017). Given that high-level automation alleviates working memory demand and older adults tend to have reduced working memory capacities, the consequences of unreliable automation may be even greater. Additionally, working memory can be taxed by increasing workload demands. This increase in workload has been shown to negatively impact task performance in younger and older adults (McBride et al., 2011). Our hypotheses were generated by drawing from these findings from the older adult literature as well as the broader automation literature (e.g., Onnasch et al., 2014; Parasuraman, Sheridan & Wickens, 2000; Endsley & Kiris, 1995).

1. Consistent with previous literature, we expect:
  - a. Older adults will exhibit lower performance compared to younger adults.
  - b. High workload will hinder performance more than low workload.



- c. Increasing degree of reliable automation will enhance performance.
  - d. Reliable automation will lead to better performance than unreliable automation.
2. We expect that compared to younger adults, older adults' performance will interact with reliability. Specifically, we anticipate that performance will decrease with unreliable automation and increase with reliable automation as degree of automation increases.
3. We expect that, compared to younger adults, older adults' performance will interact with workload and degree of automation. Specifically, we anticipate that performance will decrease as degree of automation increases under high workload compared to low workload.

## METHOD

### **Participants**

Forty three community-dwelling older adults (22 females;  $M_{\text{age}} = 72.2$ ,  $SD = 3.37$ ) were recruited and compensated \$20 for their time. Forty four college students (29 females;  $M_{\text{age}} = 18.8$ ,  $SD = 1.39$ ) were recruited from the Clemson University participant pool and compensated with course credit in exchange for their participation. Data from 4 participants were excluded from analyses (3 older adults and 1 younger adult) because their performance was lower than the requisite score of 85% on the math portion of the working memory task.

## Materials

*Equipment.* PC-compatible (Windows 7) computers running at 3.2 GHz with 4GB of RAM was used with a 19-inch LCD monitor set at a resolution of 1024 x 1280 pixels. Participants sat approximately 18 inches from the monitor, using a mouse (on the preferred side) and a keyboard.

*Taxi dispatching task.* The task was adapted from previous studies (Rovira et al., 2017; Rovira, McGarry, and Parasuraman 2007; Pak et al., 2017). The task screen was split into four parts: a street map with a grid overlay (right), a target input area which contained automated assistance (left), and a communication module (upper-left). The map display depicted customers (green boxes C1 to C6 for high workload; C1 to C3 for low workload), taxis (red boxes T1 to T6 for high workload; T1 to T3 for low workload), one headquarters (orange box labeled HQ) and three extraneous boxes (yellow boxes B1 to B3). The primary task was to observe the map and dispatch the taxi/customer pairing which were closest in proximity to one another. If two sets of customers and taxis were equidistant from each other, the pair closest to HQ took priority.

Participants dispatched taxis by selecting a customer and taxi from the target input area. To assist participants in the task, three conditions were created: a lower level information analysis automation aid (Figure 1), a medium-level decision automation aid (Figure 2), and a high-level decision automation aid (Figure 3). For the information analysis automation condition, a dispatching selection chart provided an unordered list of the distances from customer to taxi and customer to HQ. The list relieved the operator from having to manually calculate the distances between each taxi and customer. However, the level was still considered low because the list of distances was unordered.

The operator was still required to visually search through all the numbers and retain the lowest value in working memory while comparing it to all other values in the list. This particular task could be effortful for the operator to complete. The medium-decision automation condition calculated the distances and provided the closest three customer/taxi pairs to the participant in an ordered list. This alleviated working memory demand by reducing the number of options and ordering them from best to worst. The medium-decision automation was considered a comparatively higher degree of automation because the task it completed was more complex and the automation took over more of the task than the information automation. The participant was not given the distances but they could choose to “view distance calculations”, “select best” or disregard the automation. The high-decision automation performed the same task as the medium-decision automation except it presented the top choice instead of the top three. This is considered the highest level of decision automation because it only provided the participant with one option. For each condition, the next trial began if the participant either made a selection or did not make a selection within the allotted amount of time.

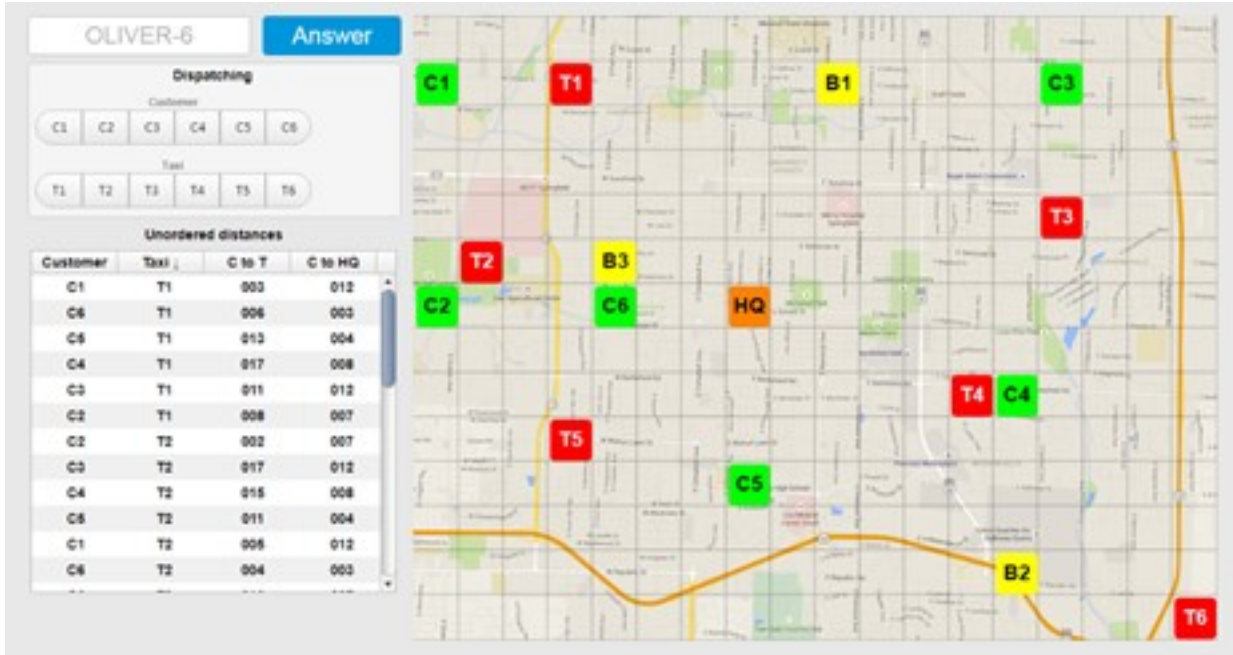


Figure 1. Information analysis automation, high workload. All taxi (T) and customer (C) distances are calculated and presented in an unordered list.

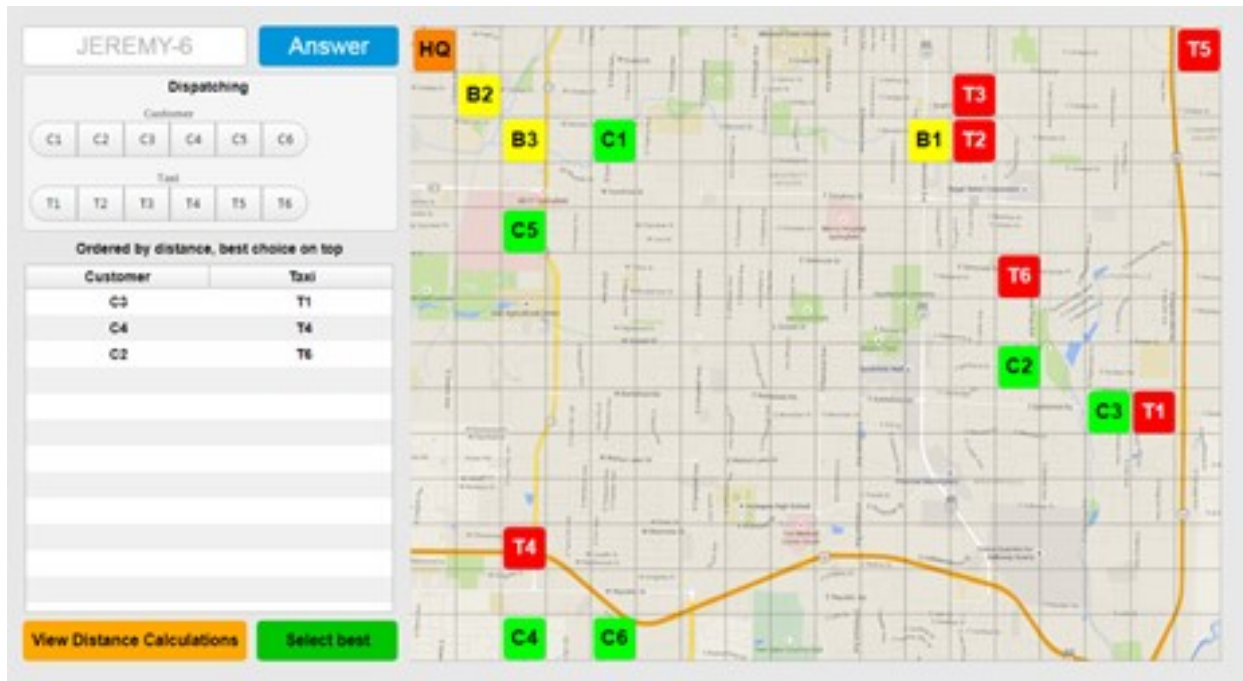


Figure 2. Medium-decision automation, high workload. All taxi to customer distances are calculated and the top 3 choices are presented to the participant.

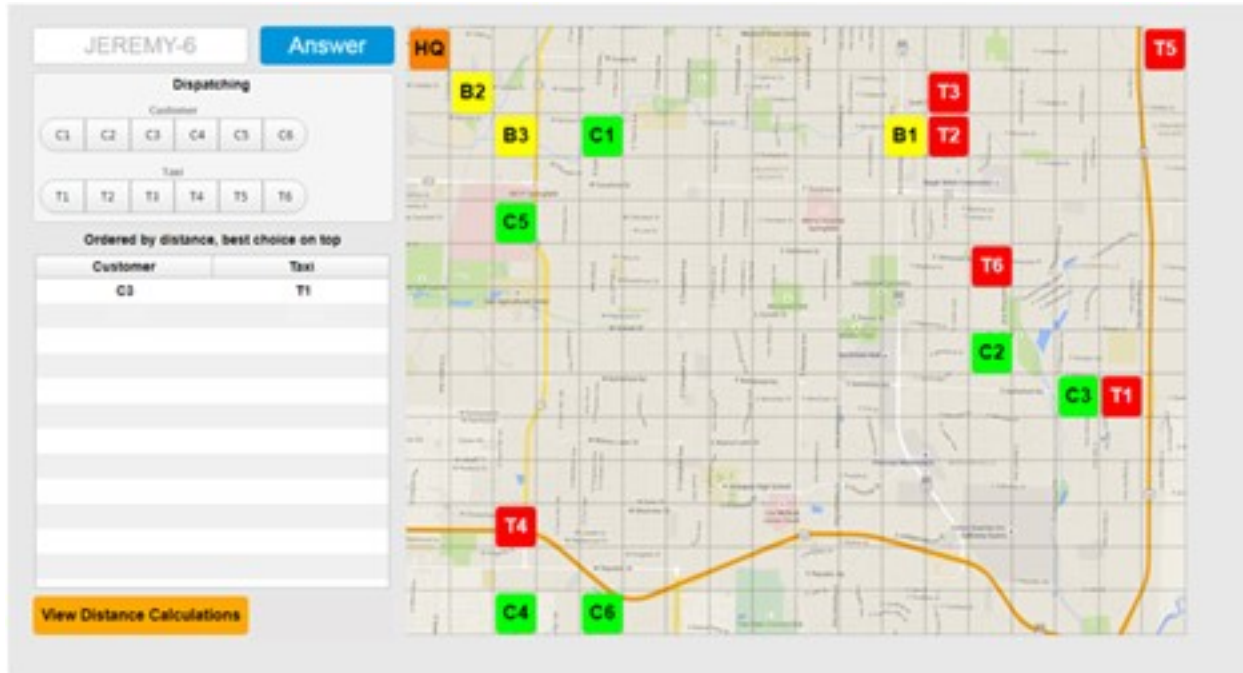


Figure 3. High-decision automation, high workload. All taxi to customer distances are calculated and the top choice is presented to the participant.

A secondary task was included because the effects of automation on performance and complacency are most often seen in multitasking situations (Parasuraman & Manzey, 2010). For this communications task, participants were instructed to monitor the communications panel (upper left) looking for a particular call sign which appeared every 6 seconds. If the specified call sign appeared, they are required to click the “ANSWER” button. This secondary task was performed during each block, there was no single task condition.

## Measures

*Working memory.* Working memory span was measured using the automated operation span task (OSPAN), a computerized version of the operation span memory task (Unsworth et al. 2005). The OSPAN was chosen because it is a highly reliable measure

and automated so there was little need for researcher intervention while participants were in the lab. The OSPAN asks participants to complete simple math problems while remembering the order of letters that are being presented to them between each problem.

*Individual differences and trust.* Preexisting individual differences in attitudes toward automation were measured with the automation-induced complacency potential (AICP) scale (Merritt et al., 2019). AICP consists of 10 items on a five-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). The primary focus is on trust in various forms of automation, including assessments of relative reliability between automated and human assistants. We also used the Complacency Potential Rating Scale (CPRS) (Mollay, Parasuraman, & Singh, 1993) as an additional, more traditional, measure of complacency potential. History-based trust was measured after each block by posing four questions adapted from Lee and Moray (1994) (see Appendix A). Participants answered using a 0-100 visual analogue scale, where higher scores indicated higher perceived trust. Subjective workload was also measured after each block using the NASA-TLX subjective workload scale (see Appendix B). This index measures subjective workload using 6 items, each assessing a different attribute of workload: perceived mental demand, perceived physical demand, perceived temporal demand, perceived effort, perceived performance, and perceived frustration. Composite scores are produced by taking a weighted average of the 6 items.

*Attentional Control.* Attentional control information was gathered as an exploratory variable from younger adults only. Attentional control was measured using a custom, PC-based version of the anti-saccade task based on Draheim, Mashburn, Martin, and Engle (2019). The main feature of this kind of attentional control task was its use of

accuracy rather than reaction time (Draheim et al. 2019). Participants were instructed to focus on a fixation point in the middle of the screen. They were then informed that a stimulus would flash on either the right or left side of the screen and a letter would appear on the opposite side. Their goal was to suppress the automatic saccade toward the flashing stimulus in order to perceive the letter, which they reported on the following screen.

### **Procedure**

The experiment used a 3 (degree of automation: information analysis, medium-decision, high-decision) x 2 (workload: low, high) x 2 (age: younger, older) mixed-factorial design. Participants completed six blocks of 50 trials for a total of 300 trials. The automation reliability was set at 80% (Wickens & Dixon, 2007), therefore, in each block, 10 trials provided incorrect assistance. For the information analysis automation conditions, the display showed incorrect distances between customers and taxis. For the medium-decision automation conditions, the best three pairings were incorrect. For the high-decision automation conditions, the single option presented to the participant was incorrect. There were no automation failures before the 10th trial to establish trust in the automation (Wickens, Hellenberg, & Xu, 2002). The subsequent automation errors were randomly distributed among the remaining trials for each block. The participant continued to the next trial once they submitted their pairing or after 13 seconds elapsed or 20 for older adults (Pak et al., 2016), whichever comes first. Each block contained either information analysis, medium-decision, or high-decision automation and blocks was randomized for each participant. Workload was manipulated by the number of pairs of taxis and customers presented on the screen, three pairs for low workload and six pairs



for high (Rovira et al., 2017). The main dependent variable was task performance, a.k.a decision accuracy, which was calculated as a ratio of how many trials each participant chose the correct answer compared to the total number of trials. Participants were run in groups of up to 5 at a time on individual computers with partitions between each person to avoid potential distractions. Participants signed the informed consent and immediately completed the working memory task outlined above. Once each participant completed these tests, they were shown instructions for using the taxi dispatching simulator. They were told that automation was present for all of the conditions but that it was not perfectly reliable. Once any questions were answered concerning the task, participants completed 12 practice trials composed of 4 information analysis, 4 medium-decision, and 4 high-decision conditions at a low workload. The participants completed the taxi dispatching task allowing for short breaks between blocks when necessary.

## RESULTS

An a priori power analysis determined that a sample of 53 participants would be required to detect a large effect size ( $f = 0.50$ ) with 90% power ( $\alpha = .05$ ). Before conducting analyses, 4 participants were excluded (3 older adults, 1 younger adult) because they scored below 85% on the math portion of the working memory test. Data from 83 participants was used for analysis. Multiple imputation was performed due to missing data. Outlier analyses were conducted but no participants were removed because their data did not appear to influence the results of the statistical analyses. Participant descriptive statistics are outlined in Table 1. Independent sample  $t$ -tests demonstrated no

significant differences in CPRS, AICP, or NASA-TLX scores between older and younger adults ( $p > .05$ ) (see Appendix C).

Table 1  
*Participant characteristics by age group including means and standard deviations for age in years, complacency potential rating scale (CPRS) ratings, and automation induced complacency potential (AICP) ratings.*

	Younger Adults				Older Adults			
	Male		Female		Male		Female	
	<i>(n = 13)</i>		<i>(n = 30)</i>		<i>(n = 17)</i>		<i>(n = 23)</i>	
	M	SD	M	SD	M	SD	M	SD
Age	19.15	1.41	18.60	1.38	72.59	3.61	71.87	3.22
CPRS	64.77	5.73	63.97	5.67	61.47	4.24	62.57	5.38
AICP	33.38	5.27	30.57	3.94	34.88	3.97	32.13	4.79

*Note.* All statistics represent data from 83 participants. CPRS scores range from 50 to 77 such that higher scores indicate higher complacency potential. AICP scores range from 21 to 45 and should be interpreted in the same fashion as CPRS.

All variables were checked for normality prior to statistical tests. The normality check revealed multiple variables violated the normality assumption. Therefore, we utilized conservative estimates for the following analyses.

### **Subjective Workload**

We measured subjective workload to help affirm that our workload manipulation was successful (i.e., low workload conditions were perceived as such than high workload conditions). Unfortunately, some data were lost due to technology errors so analyses were conducted on smaller sample sets. We expected that subjective workload ratings would be higher for high workload conditions and lower for low workload conditions. A repeated measures ANOVA revealed no main effects of workload or automation (Appendix D). We expect that differences may not have manifested in these ratings

because of automation use. Specifically, subjective workload ratings may have been lower if automation helped relieve workload demands when completing the task.

### **Trust and Automation**

We did not have any specific hypotheses concerning trust, however, we were interested in exploring whether there were differences in trust across workload or automation conditions. Analyses showed there was a main effect of degree of automation ( $F(2,81) = 26.26, p < .001, \eta_p^2 = .249$ ) but not a main effect of workload ( $F(2,81) = 12.89, p < .001, \eta_p^2 = .241$ ). The interaction between the two variables was also significant ( $F(2,81) = 6.55, p < .001, \eta_p^2 = .241$ ). Post-hoc analyses showed that trust was significantly lower for information analysis ( $M = 40.14, SD = 15.21$ ) conditions compared to medium-decision conditions ( $M = 47.70, SD = 15.65$ ); ( $t(82) = 5.87, p < .001$ ). Trust was also significantly lower under information analysis conditions compared to high-decision conditions ( $M = 48.39, SD = 17.01$ ); ( $t(82) = 5.66, p < 0.001$ ). These results are shown below in Figure 4.

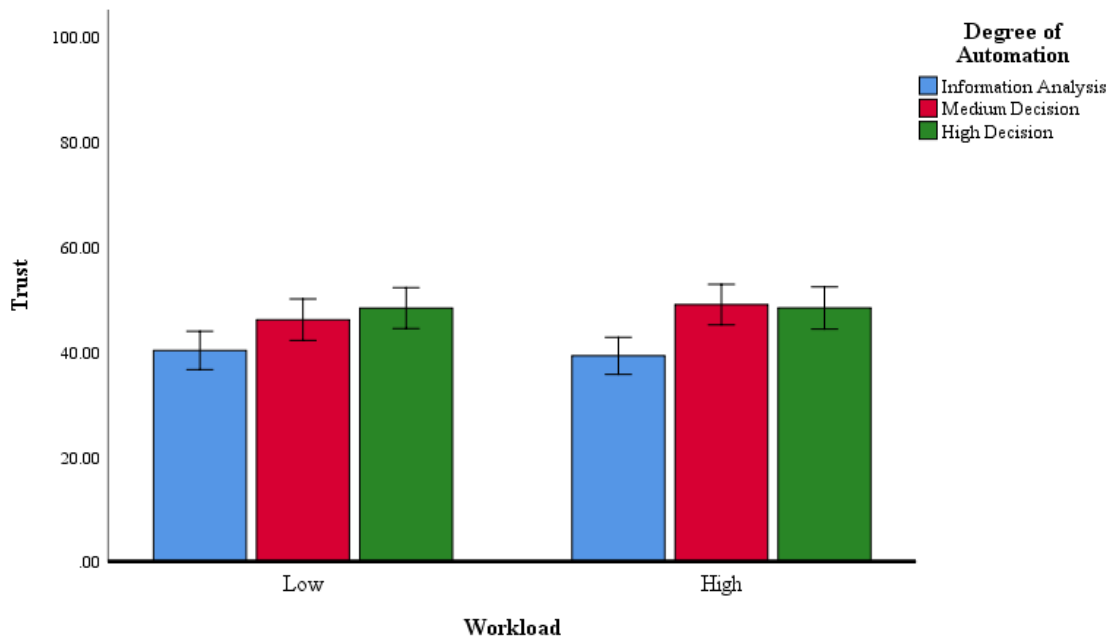


Figure 4. Trust as a function of workload across degrees of automation. Participants answered using a 0-100 visual analogue scale, where higher scores indicated higher perceived trust.

### Working Memory Differences and Age

To determine whether there were age-related differences in working memory capacity, we compared working memory scores between the younger and older adults. Working memory capacity was measured using the OSPAN task. Scores were calculated by first identifying each set of perfectly recalled letter strings and summing the number of letters in those strings producing scores between 0 and 75. We performed outlier analyses and no scores were excluded. Consistent with the literature, there was a significant effect of age group on working memory score, ( $t(83) = 2.74, p = .008$ ), with younger adults ( $M = 33.70, SD = 19.15$ ) demonstrating higher working memory capacity scores compared to older adults ( $M = 23.00, SD = 16.35$ ) (see Appendix E). The following analysis compares performance between younger and older adults to determine if older adults exhibit a

greater performance decrement with unreliable automation compared to younger adults (i.e. the lumberjack effect).

To investigate the role of working memory in task performance, we correlated working memory capacity with performance on the main taxi dispatching task across all workload and automation conditions. Analyses were conducted separately for younger and older adults. For younger adults, there were significant correlations between working memory and performance in high workload ( $r(42) = .317, p = .038$ ), medium-decision ( $r(42) = .304, p = .048$ ) and high-decision ( $r(42) = .342, p = .025$ ) conditions. For older adults, there were no significant correlations between working memory and task performance. This indicates that for younger adults, performance was better for those with higher working memory in certain conditions. However, for older adults, performance did not appear to have a relationship with working

### **Task Time Between Age Groups**

We measured task time in order to see whether there were completion time differences between older and younger adults. Additionally, we wanted to investigate whether the task time took longer depending on workload or automation condition. There was a main effect of workload ( $F(2,79) = 30.38, p < 0.001, \eta_p^2 = 0.278$ ) but there was not a main effect of degree of automation ( $p > .05$ ). The interaction between workload and age group was also not significant. On average, younger adults took significantly less time than older adults to complete the task ( $t(81) = 13.46, p < .001$ ). (see Appendix F). Additional analyses of task time are illustrated in Appendix H.

## Secondary Task Performance

Analyses were first conducted to ensure that performance on the secondary task was not significantly different across workload or automation conditions. Differences could indicate a speed/accuracy tradeoff. A repeated measures ANOVA was conducted and found a main effect for automation ( $F(2,72) = 6.46, p < 0.001, \eta_p^2 = 0.306$ ) but not for workload ( $F(1,73) = .137, p = .712, \eta_p^2 = .002$ ). Post-hoc analyses showed that performance was significantly lower during medium-decision automation conditions ( $M = .53, SD = .28$ ) compared to low automation conditions ( $M = .59, SD = .28$ ); ( $t(86) = 3.12, p = .002$ ). Performance was also significantly higher for high-decision conditions ( $M = .60, SD = .29$ ) compared to medium-decision conditions ( $M = .53, SD = .28$ ); ( $t(86) = 5.337, p < .001$ ). Despite these statistical differences, there does not appear to be a drastic tradeoff between speed and accuracy based on secondary task performance.

## Taxi Dispatching Task and Decision Accuracy

To investigate whether there was a difference in performance between age groups (hypothesis 1), we conducted a 2 (age group: younger or older) x 3 (degree of automation: information or medium-decision or high-decision) x 2 (workload: low or high) x 2 (reliability: reliable or unreliable) repeated measures ANOVA. There was not a significant interaction between age group and any of the other three variables so the following effects summarize data for all participants. There was a main effect for workload (hypothesis 1a) ( $F(2,80) = 103.48, p < .001, \eta_p^2 = .561$ ), degree of automation (hypothesis 1b) ( $F(2,80) = 8.00, p = .001, \eta_p^2 = .167$ ), and reliability (hypothesis 1c) ( $F(1,81) = 36.04, p < .001, \eta_p^2 = .308$ ). These main effects were qualified by a significant 3-

way interaction between DOA, workload, and reliability ( $F(2,80) = 32.72, p < .001, \eta_p^2 = .450$ ).

The source of the 3-way interaction, illustrated in Figures 5 and 6, was a significant 2-way interaction between reliability and degree of automation within the low workload conditions ( $F(2,81) = 30.84, p < .001, \eta_p^2 = .432$ ) but not within high workload conditions (hypothesis 2). There was not a significant interaction between workload and degree of automation (hypothesis 3) ( $p > .05$ ). For low workload conditions, when automation was reliable, individuals' performance was higher during medium-decision automation trials ( $M = .78, SD = .10$ ) compared to information analysis automation trials ( $M = .692, SD = .164$ ), ( $t(82) = 4.61, p < .001$ ) and even higher during high-decision automation trials ( $M = .875, SD = .009$ ), ( $t(82) = 8.72, p < .001$ ). That is, performance significantly increased with each increasing degree of automation when the aid was reliable. When automation was unreliable, however, performance was significantly worse during high-decision automation trials ( $M = .643, SD = .029$ ), ( $t(82) = 4.37, p < .001$ ) compared to medium-decision ( $M = .75, SD = .26$ ) and information analysis trials ( $M = .73, SD = .19$ ), ( $t(82) = 3.04, p < .001$ ).

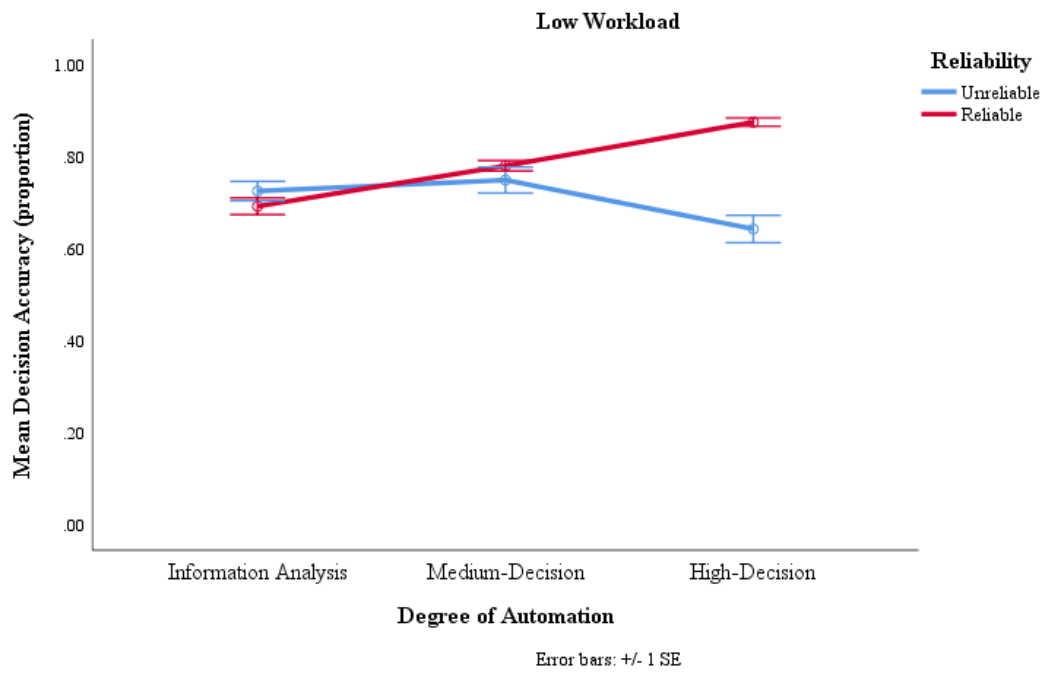
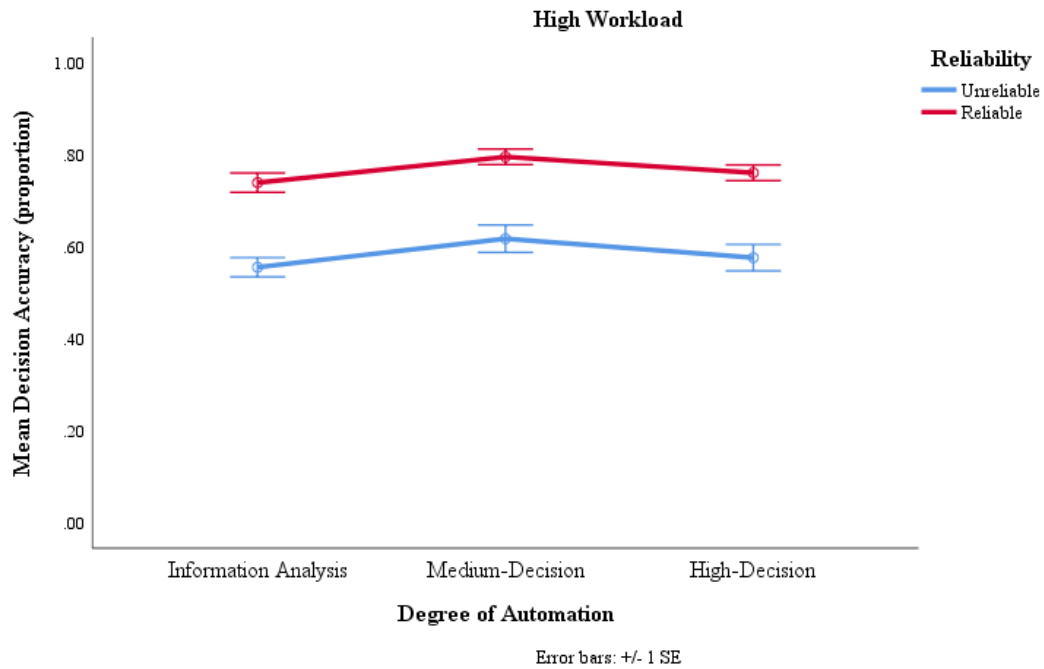


Figure 5. Mean decision accuracy as a function of degree of automation and reliability for low workload conditions as a proportion. Error bars display +/- 1 standard error.





*Figure 6.* Mean decision accuracy as a function of degree of automation and reliability for high workload conditions as a proportion. Error bars display +/- 1 standard error.

A lumberjack effect would appear as a significant increase in performance with reliable automation and a significant performance decrease with unreliable automation across increasing degrees of automation.

## DISCUSSION

The purpose of the current study was to investigate how the lumberjack effect manifested in younger adults compared to older adults. Hypothesis 1a posited that older adults' performance would be lower compared to younger adults. This was not supported, there was no main effect of age. Next, we hypothesized that performance would be better

under low workload conditions compared to high workload conditions (1b). This hypothesis was supported. Hypothesis 1c expected that performance would improve across increasing degree of reliable automation. This hypothesis was supported. Finally, we expected that reliable automation would lead to better performance compared to unreliable automation. This hypothesis was supported. To summarize these initial findings, there was not a main effect of age but there was a significant main effect of workload, reliability, and degree of automation.

Next, we hypothesized that 2) older adults would exhibit a greater lumberjack effect compared to younger adults. That is, we expected that performance would significantly decrease with unreliable automation and increase with reliable automation and this effect would be stronger for older adults. Though we saw a lumberjack effect for both age groups, there was not an increased effect for older adults. Therefore, our hypothesis was not supported. This finding contributes to the literature by demonstrating a lumberjack effect in older adults which has not previously been shown. While our older adults did indeed have reduced working memory compared to younger adults, they were not more detrimentally affected. This suggests that the source of the lumberjack effect observed in other studies may have been caused by a factor other than working memory. This finding warrants future investigation into cognitive sources of the lumberjack effect.

Finally, we hypothesized that 3) compared to younger adults, older adults' performance would decrease with increasing degree of automation under high workload compared to low workload. Again, there was no main effect of age and age group did not significantly interact with any other variable. We found that performance in high workload

conditions was not impacted more than low workload conditions by high-degree automation. This hypothesis was not supported.

The most surprising finding of this study was that there was not a significant difference in performance between younger and older adults. We expect that this was due to the specific cognitive demands of the task. For example, McBride et al. (2011) demonstrated performance differences between younger and older adults while completing two concurrent tasks, one of which utilized automation. The first task required participants to conduct a visual search under time pressure, unaided by automation. They completed a concurrent task which involved alerting automation which is a low level automation and thus requires close operator monitoring. Since this task involved a more simplistic form of automation, cognitive demands placed on participants was much higher than the task used in the current study. Thus, we would not expect to see a significant performance decrement associated with working memory capacity (i.e., normative age-related changes in cognition). Additionally, we saw that older adults on average took a significantly longer time complete the taxi dispatching task compared to younger adults (see Appendix H). This increase in response time could have helped close the performance gap between the two age groups, sacrificing speed for accuracy. Finally, we also observed lower performance on the secondary task for older adults compared to younger adults (see Appendix J). This lowered monitoring could have also played a role in helping older adults close the performance gap with younger adults. That is, reduced monitoring of the secondary task would allow for more cognitive resources to be allocated to the taxi dispatching task, thus potentially improving performance.

This study demonstrated a lumberjack effect in the lower workload conditions (increase in performance with reliable automation along with a performance decrement with unreliable automation). However, we did not observe a change in performance for the high workload conditions from the lowest to highest DOA. The finding that performance did not change under high workload conditions across reliability conditions does not support the findings in Rovira et al., (2017). They found a performance decrement under high workload conditions using unreliable automation. We suspect that these discrepancies were due to differences in our sample populations. Their sample involved cadets performing a similar task to ours using a UAV simulator. Participants may have been more comfortable completing the task in that particular domain because of their experience at a military institution. Our findings showed no significant differences in performance across DOA indicating that the high workload made the task too difficult for our sample of participants. Pak et al., (2016) did not observe a lumberjack effect under the same level of workload. However, we suspect that we were able to observe a lumberjack effect because we used a higher degree of automation which is necessary to induce OOTL effects (Parasuraman, Sheridan & Wickens, 2000). In summary, implementing higher workload conditions did not have the intended effect on performance but the addition of a higher DOA condition did produce a lumberjack effect.

Previous literature has noted the critical boundary as degree of automation moves from information acquisition and information analysis to decision selection (Onnasch et al., 2014). This is the point where consequences from OOTL effects begin to manifest (Parasuraman, Sheridan & Wickens, 2000; Endsley & Kiris, 1995). The current study adds to the literature by demonstrating a gradual decrease in performance once the

boundary point is crossed. That is, we did not observe a performance decrement from unreliable information analysis to medium-decision automation but the effect was observed when comparing the high-decision condition. This adds to the body of empirical evidence which demonstrates a performance along with a workload and situation awareness tradeoff (e.g., Endsley & Kiris, 1995; Onnasch et al., 2014; Pak et al., 2016; Rovira et al., 2017) (Appendix G). That is, as degree of automation increases so does loss of situation awareness. Operator workload also decreases as the automation adopts more of the task. There lies a theoretical point at which this tradeoff is no longer beneficial to the operator and task performance suffers. Research on these relationships should continue across domains and types of automation technologies.

### **Limitations and Future Directions**

One major limitation of this study was that there was no control condition present. It could have been informative to have a condition with no automated assistant present so that participants would have to complete the task with only manual control. Many studies include a manual control condition in order to understand the costs and benefits of implementing automation (Onnasch et al., 2014). Future studies investigating performance with unreliable automation should implement this methodology.

This study did not demonstrate a relationship between working memory and task performance despite recent literature supporting the connection between higher working memory and better task performance in older adults (Rovira et al., 2017). This may be explained by the unique cognitive demands of this particular task. The need for information maintenance in the presence of interference is viewed as the key link

between working memory and high-order cognitive ability (Engle & Krane, 2002). Though the task in this study required a certain level of information maintenance and included interference (a secondary task), the demand may not have been sufficient to demonstrate working memory differences. Average performance scores on the secondary task were at most, .59, meaning that on average, participants did not interact with the secondary task more than 59% of the time for those conditions. Therefore, if the interference was not intrusive enough, we would not expect working memory to necessarily be predictive of performance.

Further research should be done to investigate the relationship between age-related changes in cognitive abilities and performance with various levels of imperfect automation.

## **Conclusion**

The results of this study reinforce previous findings about the detrimental effects of using imperfect automation on performance (Parasuraman, Sheridan & Wickens, 2000; Endsley & Kiris, 1995; Onnasch et al., 2013; Rovira et al., 2017). This study is the first to demonstrate a lumberjack effect in older adults. Interestingly, we did not observe degraded performance in older adults. However, workload seemed to have a significant effect on performance across age groups. This study emphasizes the effect that increased workload can have on an operator's performance. Automation has been a driving force for innovation in many industries (e.g., aviation, healthcare, transportation) but implemented poorly, can play a part in catastrophe (e.g., Boeing 737 MAX). Operator

workload should be a critical consideration when implementing automated systems across any domain.

## APPENDICES



Appendix A

**Please answer the questions below about the computer aid (automation) by clicking on the scale:**

<p>To what extent did you trust (i.e. believe in the accuracy of) the automation aid in this scenario?</p> <p>Not at all <span style="float: right;">Extremely</span></p> <p><input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/></p>	<p>To what extent did you rely on (i.e. actually use) the automation aid in this scenario?</p> <p>Not at all <span style="float: right;">Extremely</span></p> <p><input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/></p>
<p>To what extent were you self-confident that you could successfully perform without the automation aid in this scenario?</p> <p>Not at all <span style="float: right;">Extremely</span></p> <p><input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/></p>	<p>To what extent do you think the automation improved your performance in this scenario compared to performance without the automation?</p> <p>Not at all <span style="float: right;">Extremely</span></p> <p><input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/></p>

Figure 7. History-based trust questionnaire adapted from Lee and Moray (1994).

## Appendix B

Regarding the task you just completed, please answer the questions below by clicking on the scale:







<p>How mentally demanding was the task?</p> <p>Low High</p> 	<p>How successful were you in accomplishing what you were asked to do?</p> <p>Not very Very</p> 
<p>How physically demanding was the task?</p> <p>Low High</p> 	<p>How hard did you have to work to accomplish your level of performance?</p> <p>Low High</p> 
<p>How hurried or rushed was the pace of the task?</p> <p>Not very Very</p> 	<p>How insecure, discouraged, irritated, stressed, and annoyed were you?</p> <p>Low High</p> 

Figure 8. NASA-TLX measure of subjective workload (Hart & Staveland, 1988).

Appendix C

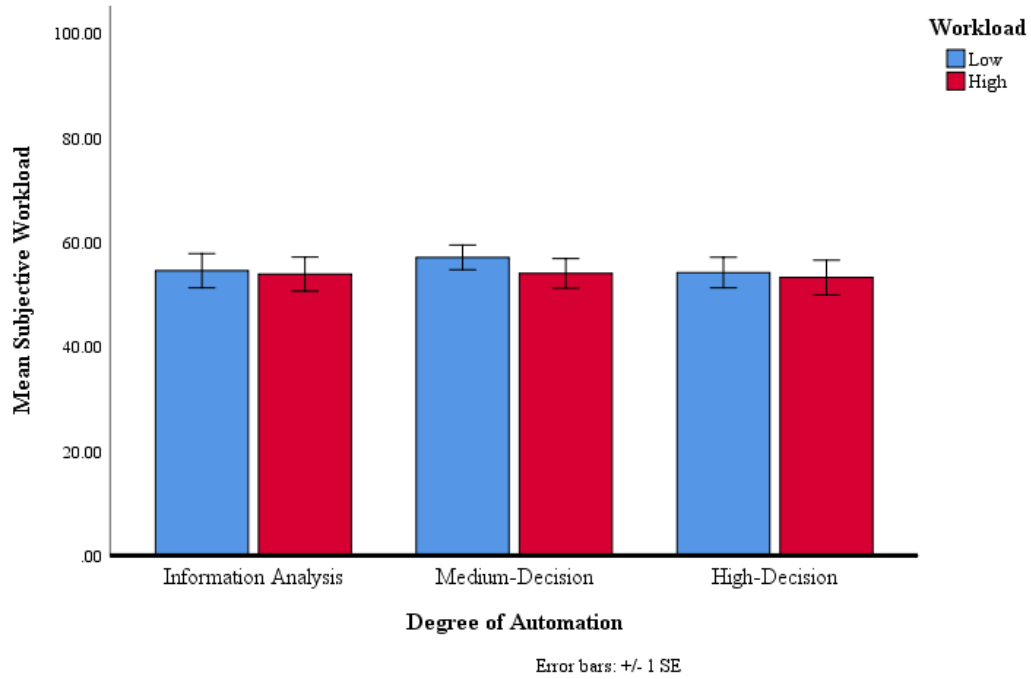
Table 2

*Mean scores for AICP, CPRS, and NASA-TLX for younger and older adults.*

	Younger Adults		Older Adults		<i>t</i>	<i>df</i>
	<i>(n = 43)</i>		<i>(n = 40)</i>			
	M	SD	M	SD		
AICP	31.42	4.51	3.30	4.61	1.88	81
CPRS	64.21	5.63	62.10	4.89	1.82	81
NASA-TLX (low workload)	50.65	11.77	54.61	10.59	1.59	79
NASA-TLX (high workload)	49.70	12.06	53.82	11.89	1.54	79

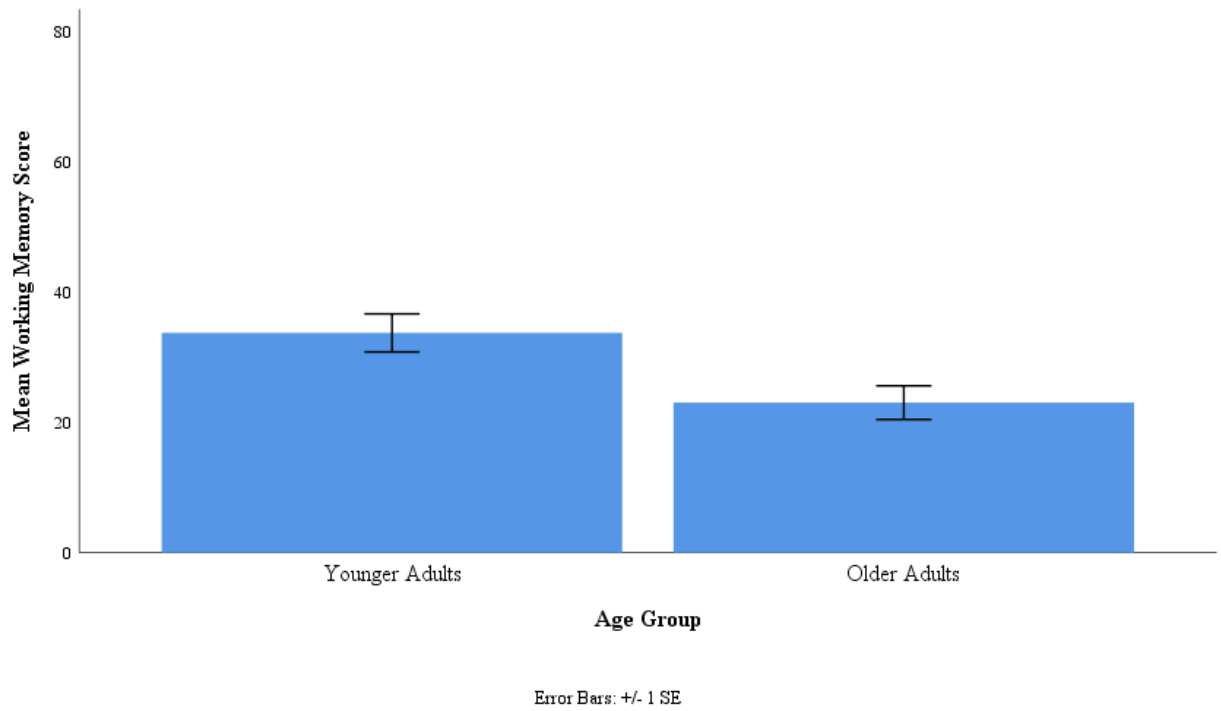
\* $p < .05$

## Appendix D



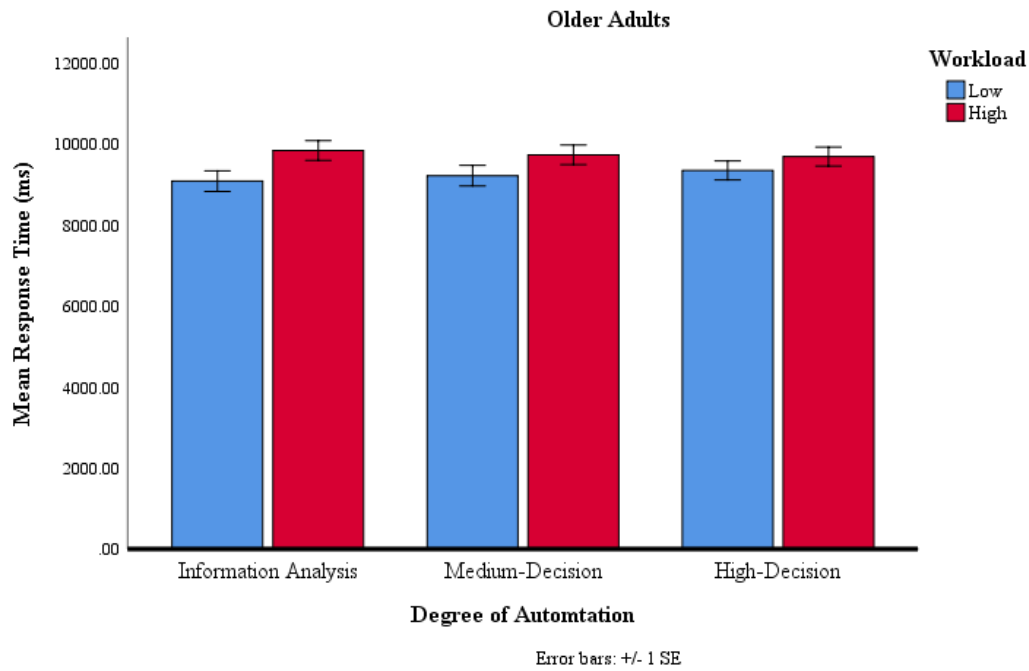
*Figure 9.* Mean ratings of subjective workload ranging from 0-100 (NASA-TLX) (Hart & Staveland, 1988). Error bars display +/- 1 standard error.

## Appendix E



*Figure 10.* Mean working memory differences between younger and older adults based on the OSPAN task (0-75). Error bars display +/- 1 standard error.

## Appendix F



*Figure 11.* Response time as a function of degree of automation and workload for older adults.

Error bars display +/- 1 standard error.

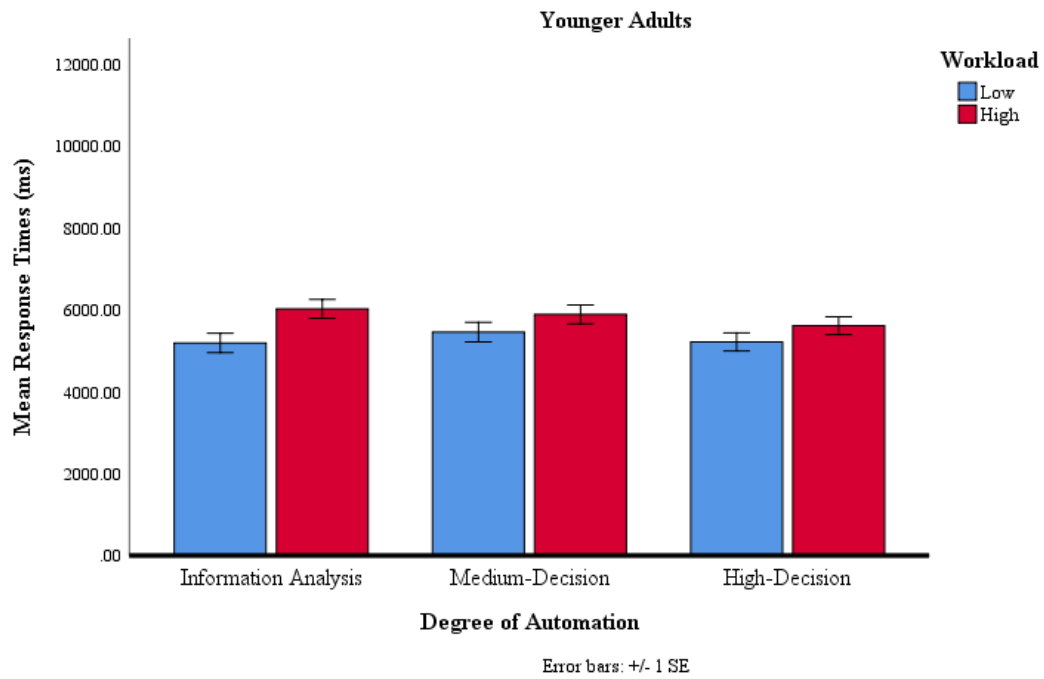


Figure 12. Response time as a function of degree of automation and workload for younger adults. Error bars display +/- 1 standard error.

Appendix G

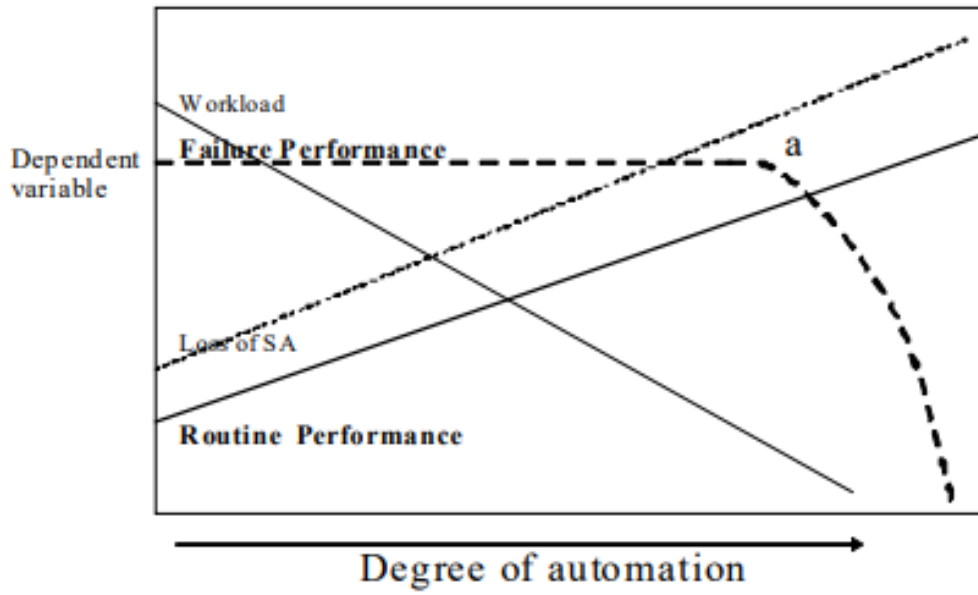


Figure 10. Tradeoff of variables, with degree of automation (Wickens et al., 2010)



## Appendix H

Analyses of task time data did not include “timed out” trials. That is, trials where participants did not provide a response were excluded from analyses. These types of trials constituted less than 5% of the original data set. To investigate the differences in task time between age groups across conditions, we first conducted a 2 (age group: younger or older) x 3 (degree of automation: information or medium-decision or high-decision) x 2 (workload: low or high) x 2 (reliability: reliable or unreliable) repeated measures ANOVA. There was no significant interaction involving age group. However, we conducted independent samples *t*-tests to determine which conditions had significant differences in task time between older and younger adults. These results are outlined in table 3. Older adults exhibited higher task times compared to younger adults in every condition. Analyses also showed that, on average, reaction times were longer for unreliable trials ( $M = 5610.5$ ,  $SD = 1243.1$ ) compared to reliable trials ( $M = 5471.4$ ,  $SD = 1019.3$ ) for younger adults, ( $t(43) = 2.38$ ,  $p = .022$ ). For older adults, reaction times were, on average, also longer for unreliable trials ( $M = 9495.0$ ,  $SD = 1587.7$ ) compared to reliable trials ( $M = 9223.3$ ,  $SD = 1314.8$ ); ( $t(40) = 2.38$ ,  $p = .010$ ). This shows that both older and younger adults, on average, took longer to decide on the optimal pairing when automation was unreliable compared to when it was reliable and older adults tended to take longer to provide an answer than younger adults.

Table 3

*Independent sample t-tests between age groups comparing response times across all conditions broken down by automation reliability (unreliable or reliable), degree of automation (IA: information analysis, MD: medium decision, or HD: high decision) and workload (low or high) in milliseconds.*

	Younger Adults		Older Adults		<i>t</i> -test
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Unreliable, IA Automation, Low Workload	5099.5	1399.4	8742.4	1828.4	-10.138**
Unreliable, IA Automation, High Workload	6276.7	1720.6	10147.6	1422.3	-11.201**
Unreliable, MD Automation, Low Workload	5630.8	1477.1	9059.6	2424.6	-7.711**
Unreliable, MD Automation, High Workload	6007.3	1717.8	9963.0	2343.0	-8.718**
Unreliable, HD Automation, Low Workload	5152.6	1571.9	9270.3	2092.2	-10.079**
Unreliable, HD Automation, High Workload	5495.8	1543.9	9787.4	2301.8	-9.901**
Reliable, IA Automation, Low Workload	5280.7	1309.4	9210.1	1877.0	-10.985**
Reliable, IA Automation, High Workload	5767.5	1473.0	8415.3	1558.4	-10.940**
Reliable, MD Automation, Low Workload	5612.1	1279.3	9157.1	1781.5	-10.346**
Reliable, MD Automation, High Workload	6090.9	1402.3	9838.2	1843.7	-10.365**
Reliable, HD Automation, Low Workload	4827.1	1129.9	8557.9	1404.9	-13.270**
Reliable, HD Automation, High Workload	5249.7	1262.3	9161.1S	1665.5	-11.991**

\* $p < .05$ . \*\*  $p < .001$

The repeated measures ANOVA also indicated a significant 3-way interaction between reliability, degree of automation, and workload. There was not a significant interaction with age so analyses were collapsed across age groups. The source of the 3-way interaction, illustrated in Appendix I, was a significant 2-way interaction between reliability and degree of automation within the low workload conditions ( $F(2,81) = 14.48, p < .001, \eta_p^2 = .263$ ) but not within high workload conditions. There was not a significant interaction between workload and degree of automation ( $p > .05$ ). For low workload conditions, when automation was reliable, mean response times were lower during high-decision automation trials ( $M = 7174.4, SD = 2540.9$ ) compared to information analysis automation trials ( $M = 7174.4, SD = 2540.9$ ), ( $t(82) = 3.07, p = .003$ ) and also compared to medium-decision automation trials ( $M = 7320.5, SD = 2350.2$ ), ( $t(82) = 4.26, p < .001$ ). That is, response times significantly decreased with high-decision automation compared to the other two conditions when automation was

reliable. When automation was unreliable, however, response times did not change significantly from information analysis trials to high-decision.

## Appendix I

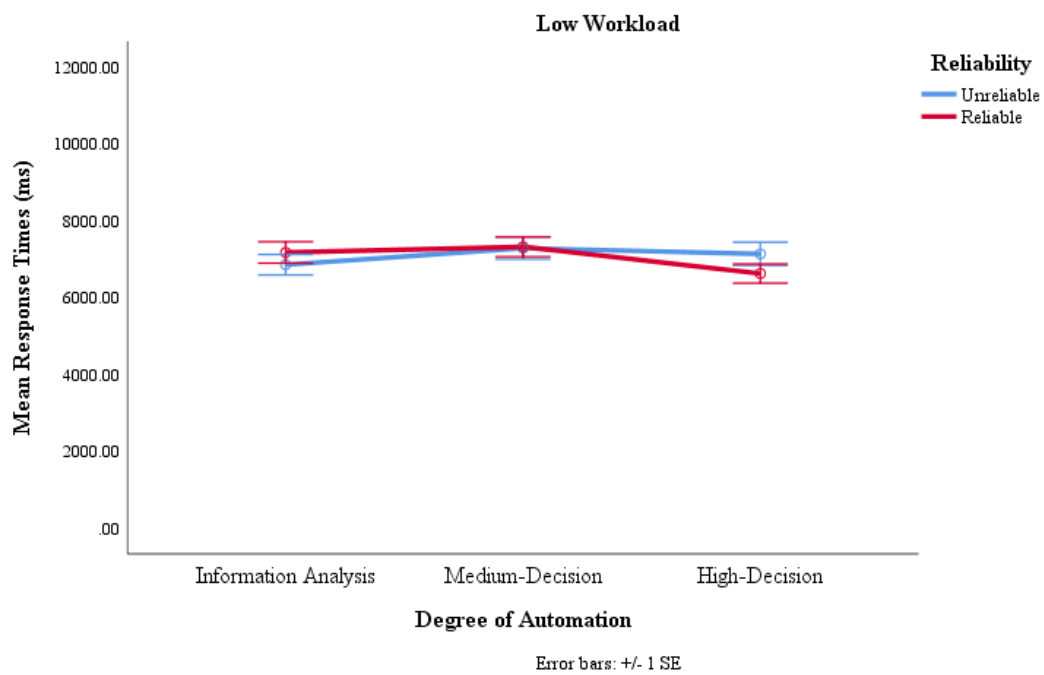


Figure 11. Mean response time as a function of degree of automation and reliability for low workload conditions as a proportion. Error bars display +/- 1 standard error.

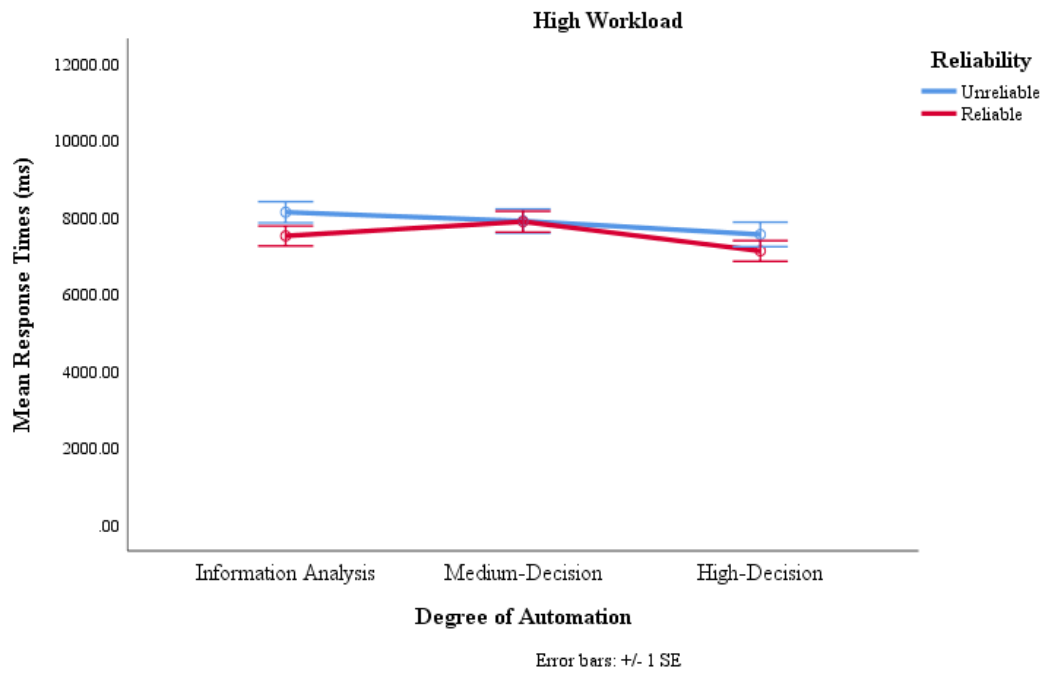


Figure 12. Mean response time as a function of degree of automation and reliability for high workload conditions as a proportion. Error bars display +/- 1 standard error.

## Appendix J

Table 4

*Independent sample t-tests between age groups comparing secondary task performance broken down by automation reliability (unreliable or reliable), degree of automation (IA: information analysis, MD: medium decision, or HD: high decision) and workload (low or high) as a proportion (0 - .1).*

	Younger Adults		Older Adults		<i>t</i> -test
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Unreliable, IA Automation, High Workload	.67	.30	.52	.34	2.09*
Unreliable, MD Automation, Low Workload	.67	.36	.49	.38	2.11*
Reliable, IA Automation, Low Workload	.66	.25	.53	.29	2.09*
Reliable, IA Automation, High Workload	.66	.24	.55	.29	2.72*
Reliable, MD Automation, High Workload	.61	.24	.47	.27	2.64*
Reliable, HD Automation, Low Workload	.64	.25	.52	.29	2.05*
Reliable, HD Automation, High Workload	.69	.23	.56	.28	2.36*

\* $p < .05$ . \*\*  $p < .001$

## REFERENCES

- Bainbridge, L. (1983). Ironies of Automation. *Automatica*, 19(6), 775-779
- 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, *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting*, 1087–1091.
- Dickson, G. (2019, March 12). Pilots warned of ‘nose down’ Boeing 737 Max 8 problems before Ethiopia crash. *Fort Worth Star-Telegram*, np. <https://www.star-telegram.com/news/business/aviation/article227481979.html>
- Dobbs, A.R., Rule, B.G. (1990). Adult age differences in working memory, *Psychology and Aging*, 4, 500– 503.
- Draheim, C., Mashburn, C.A., Martin, J.D., Engle, R.W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *American Psychological Association Psychological Bulletin*, 145(5), 508-535.
- Endsley, M. R., and Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation, *Human Factors* 37(2), 387–394.
- Hart, S.G., Staveland, L.E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Advances in Psychology*, 52, 3-4.
- Kane, M.J., Engle, R.W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review* 9, 637-671.

- McBride, S. E., Rogers, W. A., & Fisk, A. D. (2011). Understanding the effect of workload on automation use for younger and older adults. *Human Factors*, 53(6), 672-686.
- Merritt, S.M., Ako-Brew, A., Bryant, W.J., Staley, A., McKenna, M., Leone, A., & Shirase, L. (2019). Automation-induced complacency potential: Development and validation of a new scale. *Frontiers in Psychology*, 10(225).
- Onnasch, L., Wickens, C.D., Li, H., & Manzey, D.H. (2014). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 56(3), 476-488.
- Pak, R., McLaughlin, A. C., Leidheiser, W., & Rovira, E. (2016). The effect of individual differences in working memory in older adults on performance with different degrees of automated technology, *Ergonomics*, 60(4), 518-532.
- Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3), 381-410.
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230-253.
- 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.
- Rovira, E., K. McGarry, and R. Parasuraman. 2007. Effects of imperfect automation on decision making in a simulated command and control task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 49(1), 76–87.



- Rovira, E., Pak, R., McLaughlin, A. (2017). Effects of individual differences in working memory on performance and trust with various degrees of automation, *Theoretical Issues in Ergonomics Science*, (18)6, 573-591.
- Rybash, J.M., Roodin, P.A., & Hoyer, W.J. (1995). *Adult development and aging* (3rd ed.). Dubuque, IA: Brown & Benchmark.
- Salthouse, T.A. (1994). The aging of working memory, *Neuropsychology*, 8, 535–543.
- Salthouse, T.A. (1996). General and specific speed mediation of adult age differences in memory, *Journal of Gerontology: Psychological Sciences*, 51B(1), 30–42.
- Wickens, C.D., & Dixon, S.R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science*, 8(3), 201-212.
- Wickens, C. D., Hellenberg, J., & Xu, X. (2002). Pilot maneuver choice and workload in free flight. *Human factors*, 44(2), 171-188.
- Wickens, C.D. 1992. *Engineering Psychology and Human performance* 2nd. Scranton, PA: Harper Collins.