

Understanding Technology Acceptance: Effects of User Expectancies on Human-Automation Interaction

Technical Report HFA-TR-09-07

Atlanta, GA: Georgia Institute of Technology
School of Psychology – Human Factors and Aging Laboratory

2009

ANDREW K. MAYER

WENDY A. ROGERS

ARTHUR D. FISK



Requests for more information may be sent to Wendy A. Rogers
(wendy@gatech.edu)

Table of Contents

Executive Summary	4
Chapter 1 – Introduction to Human-Automation Research	7
<i>Measuring Automation Usage: Dependence, Reliance, and Compliance</i>	8
<i>Automation Error Type and User Expectations</i>	9
<i>Effects of user expectations on human-automation interactions</i>	13
<i>Summary of the Literature</i>	14
<i>Research Overview</i>	15
Chapter 2 – Study 1: Manipulating User Expectations	16
METHOD	16
<i>Participants</i>	16
<i>Design</i>	16
<i>Materials</i>	16
<i>Procedure</i>	17
RESULTS	17
<i>Participant Conceptions of Automation</i>	17
<i>Baseline Expectations</i>	18
<i>Expectation Manipulation</i>	19
DISCUSSION.....	20
Chapter 3 – Study 2: Preliminary Research on User Expectations, Error Type, and Automation Use	22
METHOD	22
<i>Participants</i>	22
<i>Design</i>	22
<i>Dependent Variables</i>	23
<i>Materials</i>	24
<i>Procedure</i>	26
RESULTS	26
<i>Ability Tests</i>	27
<i>Expectancy Questionnaire</i>	27
DISCUSSION.....	32
Chapter 4 – Summary: Study 1 & Study 2	34
Chapter 5 – Study 3: User Expectations, Error Type, and Automation Use	35
METHOD	35
<i>Participants</i>	35
<i>Simulated Scenario</i>	35
Receiving packages task	36
Dispatching trucks task.....	37
<i>Point Scheme</i>	39
Receiving packages task	39
Dispatching trucks task.....	39
<i>Materials</i>	40
Demographics.	40
Ability tests	40

General trust and specific trust questionnaires	40
Expectancy descriptions.....	40
Expectancy questionnaire	42
<i>Study Design</i>	43
<i>Procedure</i>	43
RESULTS	45
<i>Overall dependence: Block 1</i>	45
<i>Reliance and Compliance during Block 1</i>	50
<i>Reliance and Compliance during Blocks 2 & 3</i>	56
<i>Reliance and Compliance during Block 4</i>	62
DISCUSSION	69
Chapter 6 – General Discussion	73
References	75
Appendix A: Automation Attitudes	80
Appendix B: Expectancy Descriptions	85
Appendix C: Expectancy Questionnaire	88
Appendix D: Warehouse Description	90

Executive Summary

The general objectives of Phase IV of the Technology Acceptance Project were to: (1) obtain insights into the external predictive validity of our model; (2) obtain insights into the reasons why decision makers decided to accept or decided not to accept a new technology; (3) develop a *Technology-Introduction Plan* for *Deere & Company*'s introduction of new technologies in the market place, (4) develop software allowing for easy customization of scales used to measure technology and user characteristics that influence the acceptance of technologies, (5) facilitate a broader transfer of the results of the Technology Acceptance Project throughout *Deere & Company*, and (6) develop a broader understanding of the mechanisms underlying the variables of the quantitative model that will guide development of communication strategies to influence the acceptance of technologies. *This report is focused on objective (6).*

The Georgia TechAccept Model (see Van Ittersum et al., 2009) was developed to predict and understanding customers' decision to accept new technologies. In addition, we were interested in continued behavioral acceptance – how systems are used once the decision to purchase has been made. Our model identified precursor variables such as knowledge, experience, facilitating conditions (e.g., instructions), and age, as important predictors of acceptance. Efforts toward objective 6 were to understand these relationships in more depth. The results of one study are reported in Wilkison, Fisk, and Rogers (under review) which was previously submitted to Deere for review. Once this paper is published it will be made available on JD Forum. The Wilkison et al. paper demonstrated that knowledge of how automated systems work is crucial for the development of optimal reliance on such systems. If insufficient explanation or training is provided, operators are likely to make more critical errors and rely on automated systems inappropriately.

The focus of the present report is on the influence of user expectations on behavioral interactions with an automated system. We investigated these variables in several studies to understand the role of automation failure type, user expectation, and whether the age of the user was a moderating factor.

Expectations and Automation

The automation of tasks has become prevalent in many areas including the domain of commercial farming. The average age of farmers in the United States is 54.3, and 61 percent of farmers are above the age of 55, a trend that is increasing (Hernandez-Peck, 2001). The goals of automation are to increase performance, safety, and efficiency on tasks; thus adults of all ages could benefit greatly from the use of automated systems (Fisk & Rogers, 2002). It is therefore imperative that we study automation with both younger AND older adult user populations.

A number of studies have investigated different types of automation errors and their effects on behavior, but there is still some debate regarding the relative detriment different errors have on behavior. In addition, only recently has an emphasis been placed on behavioral measurements that distinguish the alarm and non-alarm state of the automation, further clouding our understanding of error type effects. Finally, although user expectations of automation performance are thought to affect behavior, little empirical research exists that attempts to systematically investigate the degree to which user expectations affect behavior when interacting with automated systems, specifically, automated alerting systems.

Three studies are presented in this report. The first two studies represent initial inquiries into the following two questions: (1) Can we subtly manipulate subjective reports of expectations of likely automation performance? (2) How do different expectations affect use of an automated system by younger adults? Study three drew on the findings from the first two studies and

endeavored to accomplish four primary goals: (1) Replicate the findings from Study 2, (2) Investigate the effects of expectations on older adult automation use, (3) Understand the time course of expectancy effects, and (4) Determine how automation error type may impact expectations and automation use in younger and older adults.

Taken as a whole, these studies make five important contributions. (1) Expectations can be easily manipulated and people are susceptible to subtle differences in automation descriptions. (2) Expectations do result in quantitatively different dependence behavior. (3) Expectancy effects can withstand some failures and successes by the automation but over time interacting with the automation expectancy apparently disappear. (4) With experience, participants draw on other sources of information to help guide behavior; in this case, information regarding errors during the alarm state of the automation guide behavior. (5) Older and younger adults show very similar patterns of behavior, although older adults tended to comply and rely on the automation to a greater degree compared to younger adults.

Chapter 1 – Introduction to Human-Automation Research

“Confident expectations of a certain intensity or quality of impression will often make us sensibly see or hear it in an object which really falls far short of it.” (James, 1890/1981, p. 402)

Automation has made many of our everyday tasks safer and more efficient (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003). Banking can be conducted faster and at our convenience, GPS navigation systems can guide us to our destination, and collision-avoidance systems warn us if we are about to collide with a hazardous object. Automation has also had an impact in the most complex systems such as nuclear power plants and airplane cockpits (Wickens & Xu, 2002). The goals of introducing automation, to increase performance, safety, and efficiency on tasks, can be accomplished by reducing operator workload and subsequent cognitive demands (Dixon, Wickens, & McCarley, 2007).

Cognitive abilities that change as individuals grow older have been well documented (e.g., Craik & Salthouse, 2000). Reaction time, visual search, divided attention, and working memory capacity have all been shown to change with age (e.g., Gottlob, 2006; Salthouse, 1992; Siedlecky, Salthouse, & Berish, 2005). These natural declines have been shown to have effects on performance of many tasks, including activities of daily living (ADLs), the result of which can lead to a loss of independence for older adults (Clark, Czaja, & Weber, 1990). However, many older adults are active in domains that require regular interaction with automated systems.

The automation of tasks has become prevalent in the domain of commercial farming. The average age of farmers in the United States is 54.3, and 61 percent of farmers are above the age of 55, a trend that is increasing (Hernandez-Peck, 2001). Given the goals of automation: to increase performance, safety, and efficiency on tasks; it seems that older adults could benefit greatly from the use of automated systems (Fisk & Rogers, 2002). It is therefore imperative that we study automation with both younger AND older adult user populations.

In both simple and complex systems, rarely is the operator/user or the automation subservient to the other, rather, they work as a “team” toward common goals (Bowers, Oser, Salas, & Cannon-Bowers, 1996). Like any team or group interaction, having an understanding of group capabilities and limitations can help gauge the potential performance of the group (Bowers et al.; Stephan, 1985). Understanding how the automated system, as a team member, is likely to perform, and the aspects of the task that it is likely to perform well, provides the operator with a basis for appropriate use (Cohen, Parasuraman, & Freeman, 1998). Appropriate use involves depending on the automation at times when it is providing correct information and disregarding the automation when it is not providing correct information. Therefore, for successful human-automation interaction the user must appropriately depend on the automation.

Measuring Automation Usage: Dependence, Reliance, and Compliance

Traditionally, researchers studying human-automation interaction have considered human usage of the automation in terms of *dependence* on the system. Dependence is a global measure of automation usage across given states of the automation (Dixon & Wickens, 2006). Recent investigations, however, have dissected user dependence into two distinct constructs: reliance on and compliance with automation (Dixon & Wickens; Meyer, 2001; 2004). Reliance occurs when the automation is silent, not asking for an action from the human. In the case of an automated warning device, reliance refers to the expected behavior of the user during the *non-alarm state*. Compliance refers to performing the action asked for by the automation. Again, considering an automated warning device, compliance refers to the expected behavior of the user during the *alarm state* of the automation.

To illustrate the distinction between compliance and reliance, imagine driving down the highway in your car equipped with an automated collision avoidance system. You begin to

change lanes, without quickly looking over your shoulder or glancing in your mirrors. The automation suddenly warns you that someone is in your blind spot. If you heed the warning and return to your lane you have *complied with* the automation. In contrast, if you begin to change lanes and the automation provides no warning and you complete the lane change, again without shoulder checking or glancing in your mirrors, you have *relied on* the automation.

The type of automation error interacts with compliance and reliance behaviors. Results indicate that behavior can be different depending on whether an imperfect automated system commits errors of incorrect rejection (misses) or errors of incorrect rejection (false alarms) (e.g., Dixon & Wickens, 2006; Dixon, Wickens, & McCarley, 2007; Meyer, 2001, 2004).

Automation Error Type and User Expectations

The type of error an automated system is likely to make depends on the response criterion. A more conservative criterion will result in the automation providing fewer false alarms at the expense of committing more misses (Wickens & Carswell, 2006). A miss occurs when the automation fails to detect the signal in the environment. In contrast, a more risky criterion will result in the automation committing fewer miss events at the expense of committing more false alarms. A false alarm is the false detection of a signal that is not present in the environment (Wickens & Carswell). Misses occur when the automation inappropriately fails to provide any warning whereas false alarms occur when the automation inappropriately provides a warning (Meyer, 2001). Because these error types are distinct and influence distinct states of interaction, appropriate interaction with automation depends on the criterion of the automation and users' understanding of that criterion.

Research suggests that both younger and older adult participants can and do adjust their behavior depending on the automation's criterion (Sanchez, 2006). Participants who interacted

with a system that only committed errors of incorrect detection (false alarms) changed their behavior such that they almost always relied on the automation but almost never complied with the automation. In contrast, participants who interacted with a system that committed errors of incorrect rejection (misses) almost never relied on the automation but almost always complied with the automation (Sanchez). Age-related differences were found in the length of time it took to adjust behavior and the degree of that adjustment. Overall, older adults were slower to adjust their behavior and when the adjustment did occur, it was to a lesser degree than younger adults.

Types of automation errors and age-related differences in dependence and performance have been investigated in select studies. Johnson (2004) tested younger and older adults using a dual-task, flight simulation where participants had to monitor engine gauges and a radar screen. The engine gauge task was aided by an 80% reliable automated system that, depending on condition, provided mostly false alarms, mostly misses, or an equal mix of both error types. Overall, older adults depended on the automation less than younger adults. In addition, within each age group, there were no statistically significant differences in dependence between the three failure conditions, although numerically, participants in the majority miss condition depended on the automation more than participants in the other two failure conditions.

Overall, younger adults outperformed older adults on the task, which was measured by combining the score on the radar task and engine monitoring task (Johnson, 2004). Within each age group, participants in the false alarm condition had the lowest overall performance, suggesting that false alarms may be more detrimental to performance than misses.

Recent evidence suggests that misses and false alarms have differential effects on reliance and compliance behavior (Meyer, 2001). An increase in misses was thought to only reduce reliance whereas an increase in false alarms was thought to only reduce compliance

(Meyer). However, a more recent study suggests that reliance and compliance may not be the independent constructs as originally proposed by Meyer (Dixon, Wickens, & McCarley, 2007). Dixon et al. found that a system with increased misses reduced reliance, but seemed to have no effect on compliance. However, they found that a system with increased false alarms reduced compliance, but also reduced reliance suggesting that false alarms have qualitatively different effects on dependence than misses. Finally, they found that false alarms had quantitatively different effects on performance, by reducing reliance and compliance, than misses.

In the study by Sanchez (2006), participants were provided with examples of potential system errors that informed participants of the difference between a false alarm and a miss. The instructions provided were standard to reliability in the sense that participants were not told what type of error would be committed by the system. This type of instruction is standard in many automation studies and, although they are neutral to reliability, they do provide users with a rudimentary understanding of system states and functioning. However, during interactions with unfamiliar automated systems, people may not have access to this type of instruction and subsequent understanding.

An important component of appropriate dependence involves processing information that provides the user with an understanding of the automation's capabilities and limitations (Cohen et al., 1998; Lee & See, 2004). There are a number of ways a user can develop such an understanding. For example, experience with the automation, which can guide future interactions with the automation, will provide the user with information regarding its reliability and performance (Chappell, 1997; Riley, 1996; Sanchez, 2006). However, with the proliferation of automated systems, individuals often face situations requiring interaction with unfamiliar systems. In such cases, users may only have expectations of likely performance to guide

dependence. Such expectations, acquired through product marketing, product documentation, related experiences, or from other users, may guide dependence behavior.

Expectancies bias information selection and often lead people to focus attention on information that confirms held expectations (e.g., Jamieson, Lydon, Stewart, & Zanna, 1987; Snyder & Frankel, 1976). In the context of automation, expectations may guide operators to select expectancy-confirming information that may guide operators' use of the system. Thus, expectancies may critically affect dependence on automated systems.

The influence of expectancies on behavior has been extensively investigated in the social cognition literature (Stephan, 1985). Expectancies bias people by priming certain information and, often unconsciously they seek out and more fully process information that is consistent with their expectancies, further strengthening those expectations (e.g., Rosenthal, 1966; Rosenthal & Jacobson, 1968). This effect of expectancies was illustrated in a study conducted by Cantor and Mischel (1977). Participants were told that they would be learning about either extraversion or introversion. Participants then read ten statements about either extraverted or introverted people. After reading the statements, participants' were presented with traits that were included or not included among the statements. Participants who expected to be learning about extraverts were more confident that they had seen non-included descriptors of extraverts compared to participants who expected to be learning about introverts (Cantor & Mischel).

Rothbart, Evans, and Fulero (1979) found they could bias participants' judgments about a group's "intelligence" by telling participants they would be interacting with a group of people who were either intelligent or friendly. If expecting interaction with an intelligent group, participants rated the group more intelligent compared to participants who expected a friendly group of people. These two studies are not isolated in terms of their findings (e.g., Feldman &

Theiss, 1982; Harris & Rosenthal, 1985; Jamieson, Lydon, Stewart, & Zanna, 1987).

Little research has evaluated expectancy as a moderator of optimal automation usage (i.e., behavioral acceptance). Expectancies, related to how an automation aide will performance, translate to *a priori* predictions of when and how well the automation will accomplish its role. Expectations can develop from prior experiences with the system or a similar system. Expectations also may develop from generally held biases or through information acquired from other users, advertisements, user manuals, or any number of other sources. The two studies reviewed next did investigate the role of expectations on human-automation interaction.

Effects of user expectations on human-automation interactions

Using an airport security luggage screening task, Madhavan and Wiegmann (2007) investigated the role of pedigree (whether the automation was labeled as an ‘expert’ or a ‘novice’) on reliance and compliance. There was no difference between the novice and expert automation other than the description of the automation that was provided to participants. During the task, participants had to indicate whether a hidden weapon (e.g., a knife) was present or absent in various x-ray luggage images. The automated aid provided participants with a determination of whether a hidden weapon was present or absent.

Madhavan and Wiegmann (2007) found that pedigree had no effect on reliance and compliance when the automated aid was 90 percent reliable, but when the aid was 70 percent reliable, participants initially relied on and complied with the expert automated aid more than the novice aid. However, after interacting with the imperfectly reliable system, participants who had the expert aid reduced their reliance and compliance. The authors proposed that the reduction in reliance and compliance was due to a violation of a ‘perfect automation schema’ suggesting that development of trust in automation is different than development of trust in humans.

Participants had an expectation of a particular level of performance that when violated by system errors caused a reduction in reliance and compliance.

Although few studies have specifically investigated the role of user expectancies, some studies have shown indirect evidence of the importance of expectations on human-automation interaction. Using a failure detection task, Chappell (1997) examined automation usage across “experienced” and “inexperienced” participants. Participants gained experience by interacting with a 100% reliable version of the system one day prior to the critical testing day. The inexperienced group interacted with the automation for the first time on the critical testing day. On the critical test day all participants interacted with a 90% reliable system. Fewer experienced participants were able to detect the first failures compared to inexperienced participants due to expectations developed during the experience-acquisition day. The previous day’s experience may have carried over into the critical test day reducing those participants’ ability to detect the initial automation failures.

Wickens, Helleberg, and Xu (2002) illustrated the influence of pilot expectations on reliance behavior. The purpose of their study was to investigate pilot maneuvering tendencies when faced with conflicting air traffic types. In that study they found that pilots who were told that the automated system was imperfect showed smaller reductions in reliance following system errors compared to pilots who were not told of the automation’s fallibility.

Summary of the Literature

The literature review brings to light a number of findings relevant to the current report. Although a number of studies have investigated different types of automation errors and their effects on behavior, there is still some debate regarding the relative detriment different errors have on behavior. In addition, only recently has an emphasis been placed on behavioral

measurements during the alarm and non-alarm state of the automation, further clouding our understanding of error type effects. Finally, although user expectations of automation performance are thought to affect behavior, little empirical research exists that attempts to systematically investigate the degree to which user expectations affect behavior interacting with automated systems, specifically, automated alerting systems.

Research Overview

Automation is prevalent in many domains. It is no longer the case that people can avoid interacting with technology with some level of automation during the average day. It is not surprising then that people may often have to interact with automated systems with which they have no prior experience of interaction. In such cases, it is likely that people rely on expectations to guide their behavioral interaction. What feeds these expectations? It could be information from a friend or a product advertisement, or expectations may be drawn from experiences with other types of automation and technology in general. It is important to understand how these vague expectations may influence automation use particularly in domains where the cost of errors is high (e.g., health care).

Three studies are presented in this report. The first two studies represent initial inquiries into the following two questions: (1) Can we subtly manipulate subjective reports of expectations of likely automation performance? (2) How do different expectations affect use of an automated system by younger adults? Study three drew on the findings from the first two studies and endeavored to accomplish four primary goals: (1) Replicate the findings from Study 2, (2) Investigate the effects of expectations on older adult automation use, (3) Understand the time course of expectancy effects, and (4) Determine how automation error type may impact expectations and automation use in younger and older adults.

Chapter 2 – Study 1: Manipulating User Expectations

The purpose of this study was three fold: (1) to discover what people think of when they hear the word ‘automation’, (2) to determine their baseline expectations of automated systems in general, and (3) to determine if these expectations can be changed through simple descriptions.

METHOD

Participants

Fifty-three students between 18 and 22 years of age ($M = 19.3$ yrs, $SD = 1.3$) were recruited from the Georgia Institute of Technology undergraduate population, twenty-five females and twenty-eight males.

Design

Participants were randomly assigned to either a high or a low expectancy group that was manipulated between participants. Twenty-six participants were presented with a high expectancy manipulation description twenty-seven were presented with a low expectancy manipulation description. The study was conducted in two stages. In Stage I, participants answered questions regarding general trust and expectancy. In Stage II, participants answered questions regarding specific trust and expectancy.

Materials

Stage I booklet. The Stage I booklet provided participants with a generic definition of automation taken from Lee (2006). Participants then completed a general trust in automation questionnaire based on Sanchez (2006) and Jian, Bisantz, and Drury (2000) and a general expectancy questionnaire that was locally developed. The final section of the booklet included open-ended questions assessing the type of automated system, if any, participants had in mind while they were completing the booklet. In addition, they were asked to list all the automated

systems they interacted with regularly and all the automated systems they were familiar with.

Stage II booklet. The Stage II booklet provided participants with a description of a specific automated system. In this case they were provided with the description of an Automated Warehouse Management System. The description was manipulated such that it created high expectations or low expectations depending on condition. The description was manipulated along four categories: Company history, technology, level of system testing, and expected performance. The descriptions were equated for both number of words and total information provided. Following the description of the system was a brief expectancy questionnaire designed to assess users' expectations of the system's likely performance.

Procedure

In groups of 2 to 6, participants provided informed consent and then completed the demographics and health questionnaire. The Stage I booklet was provided along with brief verbal instructions. Groups received either the high or the low expectancy description. When all participants had completed the Stage I booklet, the experimenter handed out the Stage II booklet. The experimenter read the expectancy description aloud to participants and participants were instructed to follow along. Participants were debriefed when everyone had completed the Stage II booklet.

RESULTS

Participant Conceptions of Automation

Participants were asked if they were thinking of a specific type of automated system when they were completing Stage I. Thirty-two participants claimed that they were not thinking of any automated system in particular when they responded to general trust and expectancy questions whereas twenty-three participants did. Figure 1 shows the categories under which

those specific systems fell and the number of times they were mentioned by participants. The *Financial* category included responses such as an ATM. Automotive included systems found in cars or trucks such as automated navigation systems. The *Home/Work Computer* category included items such as desktops and printers. The *Communication* category included items such as automated answering services. *Factory/Production* included items for mass production. The *Security* category included responses such as alarm systems. Finally, the *Other* category included anything that could not be categorized into any of the aforementioned categories.

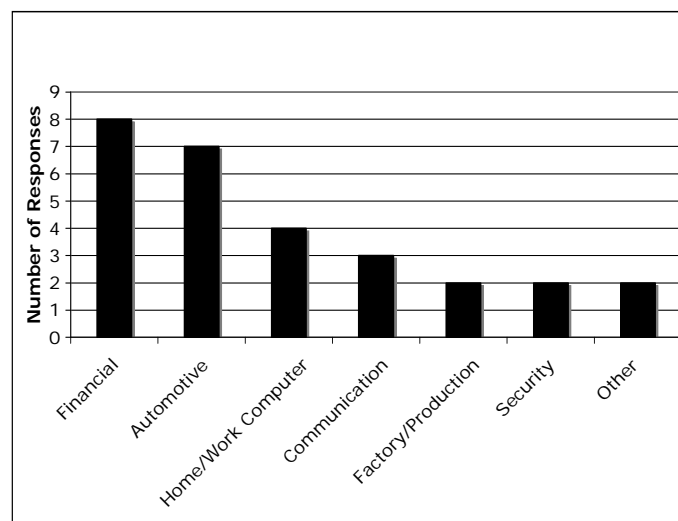


Figure 1. Categories of responses to the question “Were you thinking of a specific automated system when you were answering the previous questions?”

Baseline Expectations

In the Stage I booklet participants were asked to indicate their expectations of the general performance of automated systems. Participants indicated their expectations by providing a percentage of times they expected an automated system to perform as designed and intended. Participants in the low expectancy group had a mean expectation of 87.18 (SD=9.27) whereas participants in the high expectation group had a mean expectation of 90.25 (SD=6.94).

Independent sample t-tests indicated that this difference was not significant, $t(51) = 1.36, p = .18$.

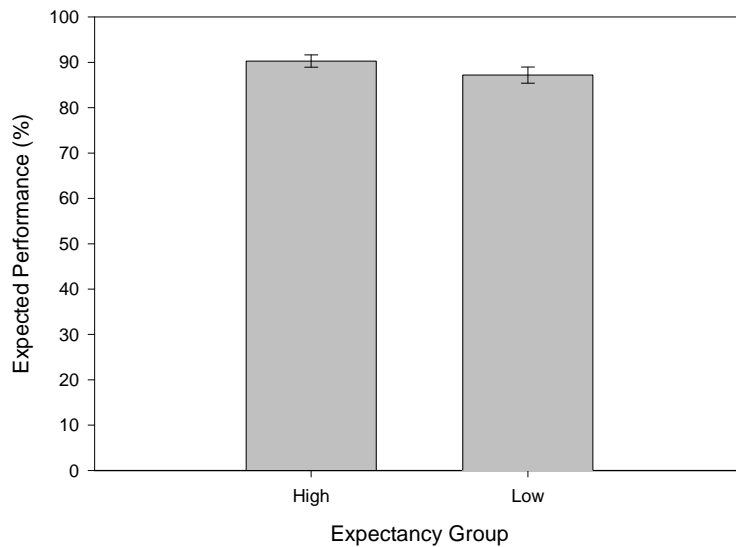


Figure 2. Baseline reported expectations by Level of Expectancy (High and Low).

Expectation Manipulation

In Stage II, participants were provided with a description of an Automated Warehouse Management System. Depending on condition, they were provided with a description that either instilled high or low expectations of likely performance. Following the description, participants were asked to provide their expectations of how well the system would perform during future use. Again participants indicated their expectations as a percentage score.

Participants in the high expectancy group expected the automation to perform at 95.17 (SD=3.93) whereas participants in the low expectancy group expected the automation to perform at 71.52 (SD=19.56). Independent sample t-tests indicate that this difference was statistically significant, $t(51) = 6.05, p < .01$ and it was a substantial difference of nearly 25%. There was also a lot more between-subjects variability in the low expectancy group; the high expectancy group was more consistent in their views.

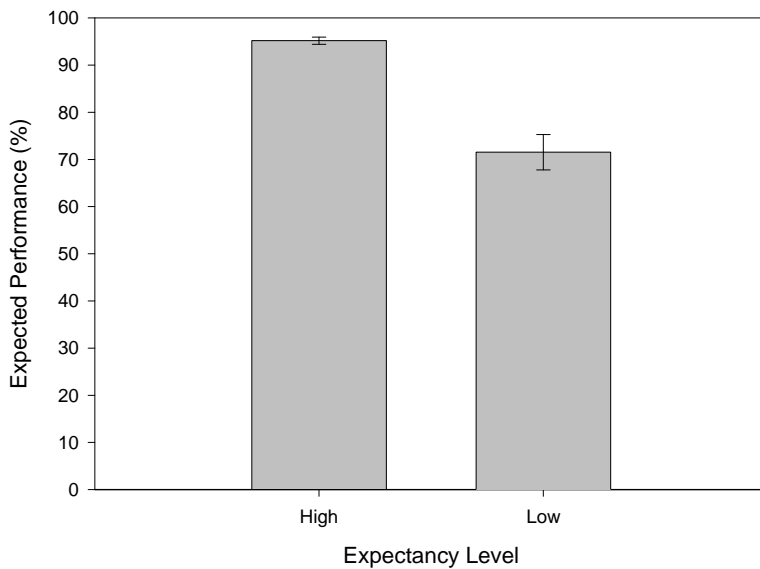


Figure 3. Reported expectancy following expectancy manipulation by Level of Expectancy (High and Low).

Paired sample t-tests were also conducted to determine if the change in expectancy from baseline to after the manipulation was statistically significant for both the low and high expectancy conditions. The nearly 5% increase in reported expectancy by participants in the high expectancy group, whose expectations changed from 90.25 to 95.17%, was statistically significant ($t(25) = 4.44, p < .05$). The over 15% decrease in reported expectancy by participants in the low expectancy group, whose expectations changed from 87.18 to 71.52%, was also statistically significant ($t(26) = 4.95, p < .05$). Thus not only did different expectancy descriptions lead to differences between the two groups, the descriptions influenced expectations within a group from before to after hearing the description.

DISCUSSION

The results from this study shed some important light on the issue of expectations of automated system performance. It is important to understand what people think about when they consider automated systems. The data from this study show that people think of systems with

which they are familiar when they answer questions related to automation performance. This finding is not surprising in and of itself but it raises the following question: What analogies are people using to make decisions when using unfamiliar automation?

Some studies in the automation literature point to an overall belief in the idea that people view automated systems as experts that never or rarely commit errors (Dzindolet, Pierce, Beck, & Dawe, 1999). In the present study, respondents anticipated automated systems to perform at approximately 89%. Although no criterion has been set to indicate what a very high or near perfect expectation in a system is, it would appear that these findings are in line with the findings from past research. But we must consider the types of systems that people were envisioning when they provided these estimates. ATMs, calculators, and in-car navigation systems are relatively robust, and if errors do occur, their results are often relatively innocuous. When people are faced with a more complex and unfamiliar system, they may inappropriately draw on experience with more common, simple automated systems. Alternatively they may trust and rely on information they receive from the manufacturer, in which case it is imperative to understand how expectations can be changed and manipulated.

We also investigated whether expectations could be manipulated using a relatively subtle manipulation. The findings show that by framing the system in terms of creating a high or low expectation we could significantly and substantially change the expectations of participants. This result is surprising given the fact that ten minutes earlier during the experiment participants provided a significantly different expectancy rating. The data point to the fragility of expectations particularly when the system is unfamiliar.

Chapter 3 – Study 2: Preliminary Research on User Expectations, Error Type, and Automation Use

The results from Study 1 illustrated the ease with which user expectations can be changed using relatively subtle descriptions. The purpose of Study 2 was to investigate how user expectations, framed through subtle descriptions, influence automation use by younger adults. Little is known about how different expectations of automation performance affect dependence and operators' abilities to detect errors. In addition, the mixed results relating to automation error types indicates a need for further research from the perspective of the state of the automation (i.e., alarm versus non-alarm state).

METHOD

Participants

Twenty-one young adults between the ages of 18 and 28 were recruited from the Georgia Institute of Technology undergraduate population. Eight females (Mean age = 20.38, SD = 1.77) and fourteen males (Mean age = 20.62, SD = 1.56) received credit in undergraduate psychology classes for their participation.

Design

The design was a 3 (level of expectation: high, low, standard) x 2 (error type: false alarm, miss) fully factorial design. Three levels of expectations were manipulated by providing participants with a description of the automated collision avoidance system with which they would be interacting. The descriptions were developed on the same principles used in Study 1 but were not identical to those used in Study 1. Those in the low expectations group heard a description that depicted the automated system as a prototype. Participants in the high expectations group heard a description that depicted the automation as an industry proven, highly reliable system. A third group, the Standard group, did not receive a description of the system

but rather were told that the automated system was highly reliable but may make errors. In addition, they were given information regarding the nature of misses and false alarms in the context of a smoke detector but were informed that these were analogous to the types of errors an automated collision avoidance system could make. Only participants in the Standard group were told about the types of errors that an automated collision avoidance system could make. The Standard group was intended to replicate the type of instructions that are often given to participants in human-automation interaction studies. In actuality, participants in all three conditions interacted with equally reliable automated systems.

Within the three levels of expectations, participants were randomly assigned to either a miss or false alarm group. Participants in the miss group interacted with an automated system that only committed incorrect rejections (misses), whereas participants in the false alarm group interacted with a system that only committed incorrect detections (false alarms). In all groups, participants were presented with a single transfer error at the end of the experimental block (i.e., an error opposite to those that participants had seen during the experiment).

Dependent Variables

Three dependent variables were measured: User expectations, non-dependence, and trust. User expectations of automation performance was measured prior to interacting with the automation using an expectancy questionnaire and acted as a manipulation check for the expectancy manipulation. Non-dependence was quantified as the number of times the participant checked the automation. Trust was measured before and after interacting with the automation using a pre and post-trust questionnaire providing a means for measuring change in subjective trust levels resulting from interaction with the automation.

Materials

Pre-test questionnaires and ability tests. Demographic and health information was collected from all participants. In addition, three ability tests, the digit symbol substitution, the reverse digit span, and the Shipley vocabulary test (Wechsler, 1997; Shipley, 1986), were administered to establish the basic memory and speed of processing abilities of the sample.

Participants completed a pre-trust questionnaire designed to evaluate their trust in a hypothetical automated collision avoidance system and to act as a baseline trust measure. Participants were then provided with the respective expectancy manipulation description followed by an expectancy questionnaire designed to evaluate participants' projections of automation performance in the upcoming task.

Stimuli. A dual-task simulation depicting a tractor equipped with an automated collision avoidance system was used for the experiment. The simulation was programmed using Flash Professional and presented to participants using a 17 inch Dell monitor powered by a Dell Optiplex GX220. The first task was a manual tracking task that required participants to keep a green square inside a bounded box. The square moved randomly in one of four directions: up, down, left, or right and participants had to use the arrow keys on the keyboard (opposite to the movement of the square) to keep it in the center of the box. If the square touched the side of the box, the square turned red. For every fifteen seconds that the square was touching the side of the box, whether it occurred at one time or accumulated over several touches, participants were penalized by losing an acre of crop. Participants were provided real-time feedback of the number of crops lost due to tracking errors.

The second task was a collision avoidance task that required participants to avoid colliding with objects located in the cornfield in front of the tractor (Sanchez, 2006). An

automated collision avoidance system was operating in the background, periodically providing warnings to participants (i.e., a red indicator with the word 'AVOID' appeared). Participants had to press the enter key to maneuver around the obstacle. When the participant appropriately maneuvered around the obstacle, an indicator with the words 'Obstacle Avoided' appeared at the top of the screen. If the participant failed to avoid the obstacle, the word 'Collision' appeared at the top of the screen and an acre of crop was lost due to the collision. If the participant maneuvered when no obstacle was present and no alarm was being presented, the words 'Unnecessary Maneuver, 1 Acre Lost' appeared. If participants pressed the enter key when a false alarm was being presented, the word COLLISION appeared at the top of the screen and an acre of crop was lost. Participants had the opportunity to check the automation by pressing and holding the space bar. This provided participants with the view of the corn field in front of the tractor. However, only one key could be pressed at any one time so the arrow keys needed for the tracking task could not be used while the space bar was being pressed. To prevent participants from over-pressing the space bar, participants were told that they would lose an acre of crop if the space bar was pressed too often but that they would not receive feedback on this aspect of the simulation.

The automation made periodic errors throughout the experimental block. The only way participants were able to detect these errors prior to a collision was by pressing the space bar to actually see or not see the object in front of the tractor. Participants in each group interacted with an equally reliable automated system. The global reliability for each group was 90% (for every 10 events 9 were automation successes and 1 was an automation failure). A total of 120 events were presented, 12 of which were error events. Each event lasted 15 seconds so the total time of the experimental block was 30 minutes.

Procedure

Participants completed an informed consent form outlining the general aspects of the study as well as their rights as participants. Upon providing informed consent, participants completed the pre-test questionnaires and abilities tests.

Participants were provided with a general definition of an automated collision system. They were then told that the automated collision avoidance system that they would be interacting with during the experiment was designed for commercial farming applications and were given a brief background on the importance of collision avoidance systems for the farming industry. Following that brief description, participants completed the pre-trust questionnaire designed to assess their trust in a hypothetical automated collision avoidance system.

The experimenter then read the expectancy manipulation description to the participant. After the verbal description, participants were given the expectancy questionnaire to complete. Complete instructions for the simulation tasks were provided to participants and they were allowed to practice the tracking task and the collision avoidance task. Participants were required to successfully conduct both tasks together error free for a minimum of three minutes before moving on to the experimental block. Following the practice block, participants completed the 30-minute experimental block.

After the experimental block, participants completed the post-trust questionnaire designed to measure their trust in the automated collision avoidance system that they had just interacted with. Finally, participants were debriefed.

RESULTS

Study 1 showed that providing participants with subtle descriptions of automated systems can have significant effects on reported expectations of likely automation performance. The

purpose of Study 2 was to determine if expectations had an effect on use of an automated system. Automation use was measured as non-dependence on the system, quantified as the number of times the participant checked the automation.

Ability Tests

See Table 1 for means and standard deviations for the ability tests.

Table 1: *Means and Standard Deviations for the Digit Symbol Substitution, Shipley Institute of Living Scale, and Reverse Digit Span.*

	Digit Symbol Substitution ¹	Shipley Institute of Living Scale ²	Reverse Digit Span ³
Mean	67.00	32.13	9.00
Standard Deviation	13.18	3.07	2.24

¹Score out of a total of 100

²Score out of a total of 40

³Score out of a total of 14

Expectancy Questionnaire

The Expectancy questionnaire was designed as a manipulation check to ensure that the automated collision avoidance system descriptions presented actually successfully framed participants' expectations. A one-way between subjects (ANOVA) was conducted with level of Expectancy (High, Low, or Standard expectations) as the between subjects factor. There was a significant main effect of expectancy ($F(2, 19) = 5.46, p < .05$). Multiple comparisons revealed that participants in the Standard group ($M = 76.89, SD = 12.53$) believed that the system would correctly detect objects significantly more often ($t(15) = 18.32, p < .05$) than the Low expectancy group ($M = 58.57, SD = 21.35$). Participants in the High expectancy group ($M = 85.67, SD = 9.42$) also believed that the system would correctly detect objects significantly more often than the Low expectancy group ($t(12) = 27.10, p < .05$). There was no significant difference between the High and Standard expectancy groups ($t(14) = 8.78, p = .29$). These results are similar to

those found in Study 1.

Non-dependence

Seven participants were excluded from the analyses due to simulator malfunction. All seven excluded participants were in the false alarm condition. The simulator failed at various times and complete data were not recorded.

An additional two participants (one from the high-miss condition and one from the low-miss condition) were omitted from the following analysis. They were excluded because their number of space bar presses far exceeded those of the other participants. Casewise diagnostics indicated that their space bar use was more than a full standard deviation from the mean and they were thus considered outliers (Figure 4 shows a scatter-plot including the outliers). It should be noted that with a larger sample these two participants may actually be a member of the normal population. However, for this study they were excluded.

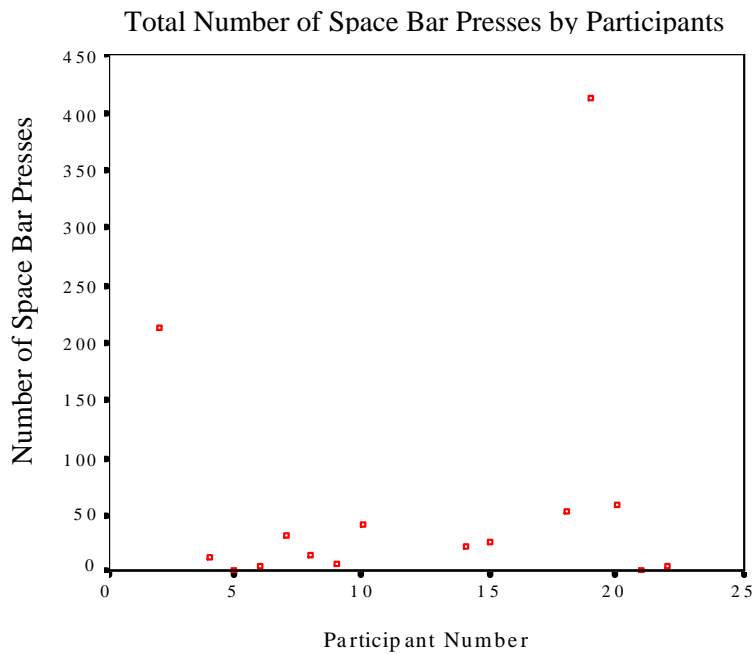


Figure 4: Total number of space bar presses by participants. Participants 2 and 19 were considered outliers.

Thirteen participants were used for the remaining analyses. The group sample sizes were: 2 in standard-miss, 1 in standard-false alarm, 2 in low-miss, 3 in low-false alarm, 2 in high-miss, and 3 in high-false alarm. Given the small sample size, emphasis will be placed on trends rather than on significance of the tests.

Overall non-dependence. Overall non-dependence was measured using the total number of times participants pressed the space bar over the course of the entire experimental block. A 2 (error type: miss or false alarm) x 3 (level of expectancy: high, low, standard) between subjects ANOVA was conducted using error type and level of expectancy as between subjects variables. There was a significant main effect for level of expectancy ($F(2, 7) = 7.81, p < .05$). Overall, participants in the high expectancy group ($M = 2.20, SD = 2.17$) depended more on the automation than the standard expectancy group ($M = 33.33, SD = 17.10$) and the low expectancy group ($M = 30.60, SD = 19.33$). Tukey multiple comparisons revealed that there were significant dependence differences between the high expectancy group and the standard expectancy group ($t(2) = 31.13, p < .05$) as well as between the high expectancy group and the low expectancy group ($t(2) = 28.40, p < .05$). Figure 5 shows the overall non-dependence behavior of participants across levels of expectancy. All levels of expectancy represent grouped behavior for participants in the miss and false alarm conditions.

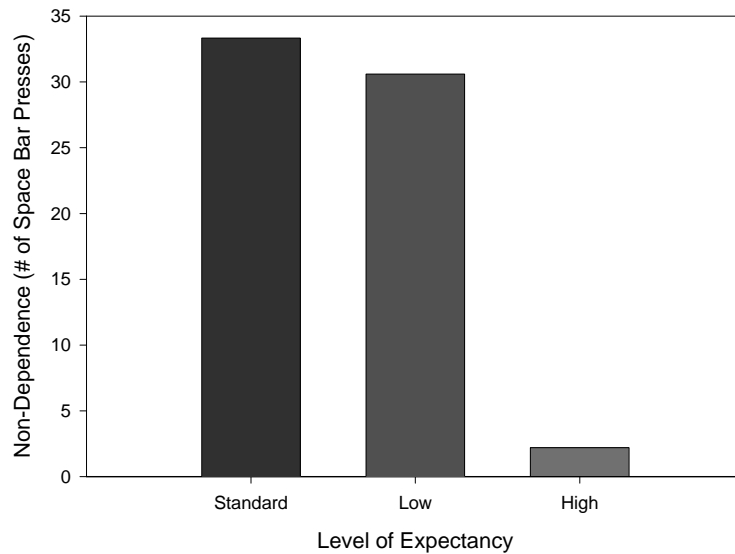


Figure 5: Overall non-reliance (number of space bar presses) by level of expectation (standard, low, or high) (n of standard group = 3, n of low group = 5, n of high group = 5).

It has been anecdotally suggested that expectancy effects last a brief period of time, up until the automation commits its first failure. To investigate this possibility, dependence data were plotted by expectancy over the course of the experimental block. Figure 6 illustrates the pattern of data. The red crosses on the x-axis represent the approximate point in time during the experimental when an automation error occurred. From the data, one can see that an error is followed by a decrease in dependence followed by a subsequent increase in dependence. However, it is clear from the data that despite automation errors, expectancy effects persist.

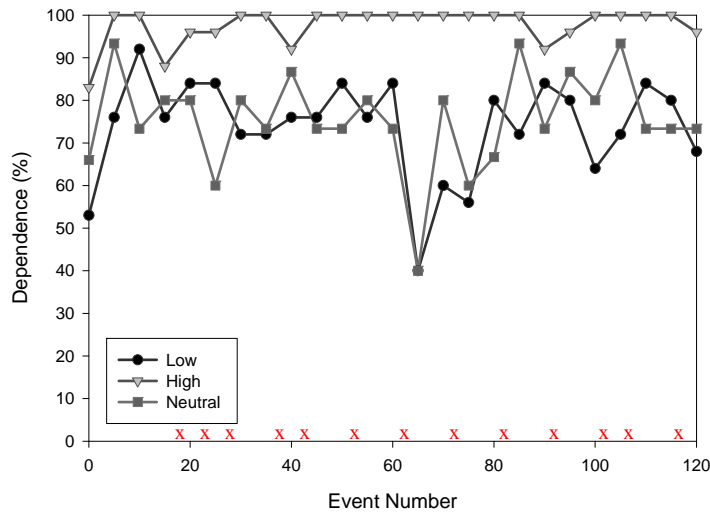


Figure 6: Dependence behavior across events during the experimental block by level of expectancy (Low, High, and Standard). The red crosses represent the approximate point in time when the automation erred.

There was no significant main effect for error type ($F(1, 7) = 3.12, p > .05$) and no significant interaction between error type and level of expectation ($F(2, 7) = 1.38, p > .05$).

However, Figure 7 suggests that, overall; false alarms elicit greater non-dependence behavior compared to misses.

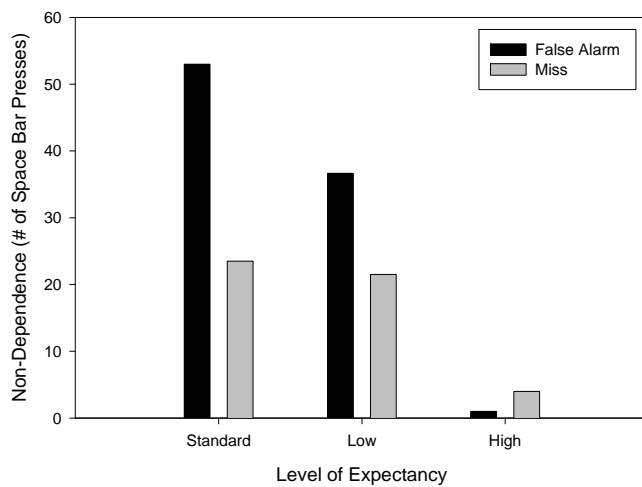


Figure 7: Overall non-dependence behavior for error type (miss or false alarm) by level of expectancy (standard, low, or high)

DISCUSSION

The purpose of this study was to investigate how manipulating participants' expectations of an automated collision avoidance systems likely performance would influence dependence on the automation. In addition, this study investigated how the types of automation errors (i.e., misses or false alarms) interact with participant expectations.

Expectancy Effects

The expectancy questionnaire was completed as a manipulation check to ensure that the descriptions significantly changed the expectations of participants and to replicate the findings from Study 1. Of particular interest were differences between the high and low expectancy conditions. Participants in the standard condition received very little information so it was difficult to predict how participants in that group would expect the automation to perform. Results indicated that participants in the high expectancy group did indeed expect the automation to perform more accurately and reliably than participants in the low expectancy condition. This finding is consistent with the findings reported in Study 1 providing support for the effectiveness of the expectancy manipulation. Interestingly, participants in the standard condition also had significantly higher expectations of automation performance compared to the low expectancy group even though only participants in the standard group heard information on the specific types of errors that an automated collision avoidance system can make. This result is unexpected, as hearing about automation errors should have primed participants to expect errors and thus should have lowered participants' expectations of the automation's performance.

Non-dependence

Study 2 was designed to make initial investigations into the effect of different user expectations on behavior when interacting with automation. These preliminary data suggest that

different user expectations significantly change the extent to which participants will depend on an automated system. Surprisingly, even in the face of automation errors, participants with high expectations continue to depend on automation at a high level. Conversely, participants with low expectations continue to depend on automation at a relatively low level despite continued automation successes. These findings suggest that expectancy effects last much longer than hypothesized by other researchers (Wickens & Xu, 2002). Further research is required to extend the period of interaction to determine the point at which expectancy effects fade, if ever.

Interestingly, participants in the standard expectancy condition depended on the automation similarly to participants in the low expectancy condition despite reporting expectations similar to participants in the high expectancy condition. One explanation is that because participants in the standard condition had some understanding of system errors, they were better able to recognize and encode automation errors, resulting in an underestimation of automation reliability. This finding has implications for human-automation interaction research in general because these types of standard instructions may prime people to expect automation errors, artificially reducing dependence behavior.

Chapter 4 – Summary: Study 1 & Study 2

Study 1 and Study 2 were designed as initial investigations into the role user expectations have on human-automation interaction. The expectancy data from both studies clearly show that subtle descriptions of an automated system can significantly change reported expectations of the automated system's likely performance. These findings have implications for training, advertising, and instructions used in automation research. However, reported expectations do not necessarily result in like behavior.

Study 2 explicitly measured the effect of user expectations on human-automation interaction behavior. The preliminary results show that expectancy effects do influence behavioral dependence on automation. In addition, the effect is sustained over time despite automation failures. The data also suggest that user expectations play a larger role in dependence than automation error type. However, limitations of Study 2 make it difficult to draw any firm conclusions. The small sample and the high level of data attrition are cause for concern. Also, the global measure of dependence limits any findings or non-findings related to automation error type. Finally, the 30 minute experimental block did not provide enough time to see changes in behavior over time. Despite the limitations, the preliminary data show that pursuing expectancy effects in human-automation interaction is a promising research endeavor.

A third study was designed to further investigate expectancy effects in human-automation interaction. The purpose was to increase the sample size and to extend findings to an older adult population. The experimental block was extended to 80 minutes and data were collected that allowed analysis at the level of dependence, reliance, and compliance. We also developed a novel simulation task that was unfamiliar to participants yet that represented a common class of automation (i.e., automating alerting system).

Chapter 5 – Study 3: User Expectations, Error Type, and Automation

Use

Study 2 provided preliminary support for pursuing research investigating the effect of user expectations and human-automation interaction. The purpose of Study 3 was to extend the findings from Study 2 by increasing the sample size, increasing the time on task, investigating age-related differences, and further studying effects of automation error type by investigating dependence, reliance, and compliance.

METHOD

Participants

Sixty older adults ($M = 70.73$, $SD = 3.59$) recruited from the Atlanta area, and sixty younger adults ($M = 20.53$, $SD = 1.69$) recruited from the Georgia Institute of Technology community participated in this study. Older adult participants were compensated thirty dollars and younger adults were compensated two and half credits to be applied to undergraduate psychology classes for their participation in the study. Participants were tested for near and far visual acuity and were required to have a minimum corrected vision of 20/40. One younger adult participant was excluded due to visual acuity below 20/40. This participant was replaced.

Simulated Scenario

A dual-task Automated Warehouse Management System (AWMS) was developed and used for the experiment. The AWMS was programmed in JAVA Script and was displayed on a 17 inch monitor connected to 3.20 GHz Pentium 4 computers. The scenario was specifically developed to meet three criteria. Firstly, the scenario needed to be novel to participants. Secondly, the scenario had to be representative of a certain class of automated system, in this case an automated alerting system that allows the user to focus more attention on a secondary

task and where checking the automation comes at a cost to the user. Finally, the scenario had to be sufficiently engaging for participants.

In the scenario, participants played the role of the shipping/receiving manager at a hypothetical warehouse. Participants were responsible for receiving shipments from incoming trucks and ensuring that departing trucks were dispatched fully loaded. The goal of each task was to earn as many points as possible. Both tasks are described below followed by the point breakdown for each task.

Receiving packages task. The general idea for this task was that the warehouse manager (played by participants) had a list of packages that were expected for delivery at the warehouse (the receiving tag list). The packages on the receiving tag list were represented by barcode-like patterns. A package with a barcode-like pattern (the shipment tag) would be delivered to the warehouse and participants had to crosscheck the shipment tag with their receiving tag list by matching the barcode-like pattern of the shipment tag with the barcode-like patterns on the receiving tag list. The barcode-like patterns were developed from a random string of five ASCII symbols

Participants matched the shipment tag to the corresponding receiving tag in the list by using the up and down arrow keys on the computer keyboard to scroll through the receiving tag list. Participants then pressed the 'receive' key to select the matching receiving tag (the 'receive' key was a keyboard key labeled with the word 'receive' in capital letters) (See Figure 8 for an example of the receiving packages task). If the correct receiving tag was selected, feedback indicating a correct response was provided, and a new shipment tag and a new receiving tag list were displayed. If the incorrect receiving tag was selected, feedback indicating an incorrect response was provided, and a new shipment tag and a new receiving tag list were displayed. If

participants took longer than the allotted time to respond, feedback indicating a time-out was provided and a new shipment tag and receiving tag list were displayed. Older adult participants were given ten and a half seconds to make a response whereas younger adult participants were given seven seconds to respond. These response times were determined based on usability testing prior to commencing this study.

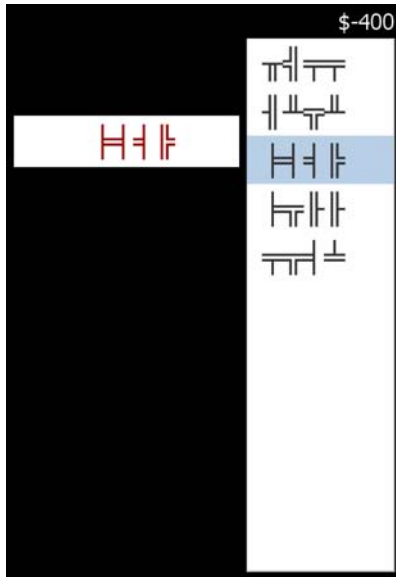


Figure 8. Example of receiving task.

Dispatching trucks task. The dispatching trucks task required participants to dispatch fully loaded trucks from the loading dock. The dispatching trucks task was supported by an automated aid (the AWMS) that, conceptually, scanned the interior of the truck trailer and estimated when the truck was full. When the AWMS estimated that the truck was full, a visual notification was provided to the participant instructing them to dispatch the truck. Participants had to then press the ‘dispatch’ key (the ‘dispatch’ key was a keyboard key labeled with the word ‘dispatch’ printed in capital letters) to notify the truck driver to leave the loading dock. Participants had the option to check the automation by manually viewing the loading truck by pressing and holding the space bar key labeled ‘view truck’ (See Figure 9 for an example of the

dispatching trucks task). However, manually viewing the truck came with a cost; that is, the receiving packages task was hidden and the arrow keys required for the receiving packages task were disconnected while the ‘view truck’ key was being pressed.

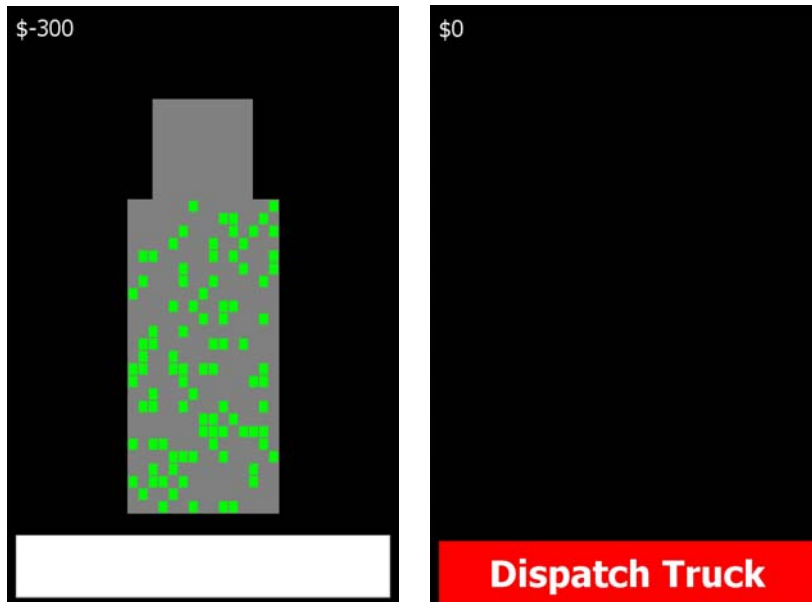


Figure 9. Left: example of truck filling as would be seen if the participant were pressing and holding the ‘view truck’ button. Right: example of the AWMS providing an automated alert indicating a full truck.

Trucks loaded at a random rate to prevent participants from estimating when a truck was fully loaded. Each truck took between 12 and 22 seconds to fill. Participants had 10 seconds to dispatch the truck once the truck was full. If the truck was not dispatched within 10 seconds of being filled, the truck overloaded and participants were penalized. If the truck was dispatched before it was full, participants were also penalized. All participants managed 160 truck loadings, divided into four 40 truck blocks, such that the total experimental session lasted approximately 80 minutes.

Participants interacted with a system that either committed false alarms (FA) or misses. In the FA condition, the AWMS periodically indicated that the truck was fully loaded when in fact it was not. In the miss condition, the AWMS periodically failed to indicate that the truck

was full when in fact the truck was full. In both the FA and miss conditions, the AWMS performed at 90 percent reliability during the first three experimental blocks and at 60 percent during the fourth block, the transfer block. Participants were not informed of what type of error the system would commit nor were they told the reliability of the system.

Point Scheme

Participants earned points for correctly receiving shipments and for dispatching fully loaded trucks. Participants lost points for incorrectly receiving shipments and for overloading trucks or sending incompletely filled trucks. Point totals were logged for the shipping and receiving tasks separately. In addition, a grand total was maintained for the combination of the shipping and receiving task points. The point tally for each task was shown throughout the experimental session. The point total for each task and the grand point total were presented at the end of each block. It should be noted that task performance, for the purpose of statistical analysis, was not measured in terms of points, but rather in terms of correct responses.

Receiving packages task. The point scheme was determined during usability testing. Older adult participants earned 15 points for correct responses whereas younger adult participants earned 10 points for correct responses. Older adult participants lost 15 points for every incorrect response while younger adults lost 10 points for every incorrect response. If participants exceeded the time limit, the system timed out and was considered an incorrect response. The faster participants performed the receiving packages task, the more shipments that could be received and the more points that could be earned. Because older adults generally required more time to make a response on the receiving packages task, they had to earn more points per correct response to equate their point performance with those of younger adults.

Dispatching trucks task. Participants received 100 points for dispatching a full truck. If

participants dispatched a truck that was not full, they lost 200 points. If participants overloaded a truck, they also lost 200 points. The reason points were differentially rewarded and penalized was because initial usability testing with younger and older adults indicated that truck dispatching errors were largely ignored due to their infrequency and low cost. Increasing the cost to 200 points for incorrectly dispatched trucks was sufficient to make participants focus on both tasks equally.

Materials

Demographics. Demographics and general health information were gathered using the demographics questionnaire.

Ability tests. Participants completed the digit symbol substitution as a measure of general perceptual speed (Wechsler, 1997), the reverse digit span as a measure of general memory span (Wechsler), and the Shipley vocabulary test as a measure of general verbal ability (Shipley, 1986). The purpose of administering these abilities tests was to ensure participants were representative of their age group and as a way to identify and exclude participants who fell outside the abilities representative of their respective age groups.

General trust and specific trust questionnaires. Prior to testing, participants completed a general trust questionnaire designed to evaluate general trust in automation. The general trust questionnaire was developed from questionnaires used by Sanchez (2006), and Jian, Bisantz, and Drury (2000) (See Appendix A for the general and specific trust questionnaires). The specific trust questionnaire was administered after participants completed the third experimental block. The specific trust questionnaire was locally developed and designed to evaluate participant trust in the automated system with which they had just interacted.

Expectancy descriptions. Expectancy was operationally defined as the participant's

belief regarding the automated system's likely performance during the experiment. Expectancy was manipulated by providing participants with a written description of the AWMS and the company that produced the system with which they would be interacting. The experimenter read the system description to participants while they followed along with a written version of the description. The description framed participants' expectations such that participants either expected good automation performance (high expectancy group) or poor automation performance (low expectancy group). A third expectancy group, the standard group, was told that the automated system was very reliable but could make errors. In addition, the standard group description provided participants with basic information regarding system misses and false alarms (See Appendix B for all three expectancy descriptions).

The high and low expectancy descriptions were manipulated along four categories: company history, type of technology, amount of system testing, and expected system performance. All other information provided in the descriptions was held constant for each manipulation. Table 1 provides a side-by-side comparison of the manipulation for the high and low expectancy groups for each category of manipulation. The standard group was designed to replicate instructions similar to those provided in previous automation studies (e.g., Sanchez et al., 2006) and was not manipulated along the same four categories as the high and low expectancy groups. In addition, the standard group provides data for comparison to findings from previous automation research.

Study 1 was designed to calibrate and test the effectiveness of the expectancy manipulations for the high and low expectancy conditions. Based on findings from Study 1, it was expected that participants in the high expectancy condition would report higher expected system performance relative to the low expectancy condition.

Table 2. Comparison of high and low expectancy descriptions

Category of Manipulation	Low Expectancy (SRT-1) In-Text Manipulation	High Expectancy (SRT-2) In-Text Manipulation
Technology	The company's first prototype system, the SRT-1, utilizes advanced decision algorithms and sensing technologies that have the ability to adjust to differing warehouse and loading conditions.	The company's latest groundbreaking system, the SRT-2, utilizes advanced decision algorithms and sensing technologies that have the ability to adjust to differing warehouse and loading conditions.
Company History	The company first became involved in sensory technologies in 2000 with the sole mission of creating advanced scanning and decision making systems for warehouse loading and shipping applications. In 2001, the company proposed an Automated Warehouse Management System called the SRT and in 2004 proposed a Smart Automated Warehouse Management System, the SRT -1.	The company first became involved in sensory technologies in 1975 with the sole mission of creating advanced scanning and decision making systems for warehouse loading and shipping applications. In 1985, the company released an Automated Warehouse Management System called the SRT and in 1997, released a Smart Automated Warehouse Management System, the SRT-1.
Level of System Testing	Testing of the SRT-1 has not begun so designers are unsure of how the accuracy, reliability, and robustness of the Automated Warehouse Shipping system will compare to the industry standard.	Testing of the SRT-2 indicates that it sets the industry standard for accuracy, reliability, and robustness and is still considered the leader in Automated Warehouse Management System systems.
Expected Performance	Because this is a first prototype Automated Warehouse Management System, it is expected that the SRT-1 will perform at a low level with some performance errors.	Because this is a well proven Automated Warehouse Management System, it is expected that the SRT-2 will perform at a high level with no performance errors

Expectancy questionnaire. Participants were provided with the respective expectancy manipulation description followed by an expectancy questionnaire designed to evaluate participants' predictions of likely automation performance in the upcoming task. The Expectancy questionnaire is presented in Appendix C.

Study Design

The level of user expectancy (low, high, or standard) and type of automation error (false alarm or miss) was manipulated as between participant variables. Exposure to the automation task was divided into blocks of time (~20 minutes per block) and were analyzed as a within participant variable. Age was a grouping variable. Each subgroup of age, younger and older adults was separated into six groups. Each group received the low, high, or standard expectancy manipulation and interacted with an automated system that either provided only false alarms or only misses. The reliability of the automated system was 90% during the first three experimental blocks. All participants were presented with a transfer block where the automation performed at 60% reliability.

Procedure

Upon receiving informed consent, participants completed the demographic and health questionnaire and the visual acuity tests.

Participants were then given a general definition of automation. Specifically, the definition read, “An automated system is a technologically-based system used to partially or fully assist the human in tasks involving sensing, detecting, information processing, making decisions and/or executing actions” (Sanchez, 2005). After reading the definition, participants completed the general-trust in automation questionnaire.

Following the general-trust questionnaire, participants were provided with general information about large-scale warehouse operations to provide context for the study (See Appendix D for Warehouse Operation description). Participants were then provided with their respective expectancy manipulation description. After receiving the description, participants completed the expectancy questionnaire. Participants had the written expectancy description in

front of them while they completed the questionnaire.

Participants were then provided with detailed instructions of the experimental task. Following the instructions, participants completed four distinct practice blocks with the simulation. The first practice block allowed participants to practice only the receiving packages task without the time limit. Younger adults had to achieve 200 points and older adults had to achieve 300 points to move on to the second practice block, meaning, at a minimum, participants had to correctly receive 20 packages. The second practice block again allowed participants to practice only the receiving packages task but with the time limit restriction. Again, younger adults had to earn 200 points and older adults had to earn 300 points to move on to the next practice block. The third practice block allowed participants to practice only dispatching trucks. Participants had the opportunity to dispatch two trucks without the aid of the AWMS. The final practice block allowed participants to practice both receiving packages and dispatching trucks simultaneously. In the final practice block participants were aided by the AWMS (that performed at 100% reliability) and consisted of five truck-loading events for younger adults and seven truck-loading events for older adults.

The experimental session consisted of four blocks, each approximately 20 minutes in duration, separated by three breaks. Each break lasted a minimum of one minute but could last longer if participants needed more time. Participants completed the specific-trust questionnaire during the third break. When the specific-trust questionnaire was completed, participants completed the final block (the transfer block). In the transfer block, the AWMS committed the same errors (FAs or misses) as it did during the experimental session but only performed at 60% reliability. Participants were not informed of the change in automation reliability. The purpose of the transfer block was to determine if participants were relying on expectancies to guide

behavior or were matching their behavior to the reliability of the system.

After the transfer block, participants completed the reverse digit span, the digit symbol substitution, and the Shipley test. Finally, participants were debriefed and compensated for their participation.

RESULTS

Four goals were set when designing Study 3. First, we sought to replicate the findings from Study 2 and to conduct the same analysis with data from older adult participants. Second, we wanted to extend the findings from Study 2 to a different task context. Third, we wanted to investigate any differential effects of automation error type. Finally, we wanted to assess the time course of effects associated with both expectancy and error type.

Overall dependence: Block 1

Dependence data from Block 1 were analyzed to determine if the results from Study 2 were successfully replicated. Block 1 was analyzed separately because the duration of Block 1 in Study 3 was similar to the duration of the experimental block in Study 2. A between subjects ANOVA was conducted using Expectancy and Error Type as between subject factors and dependence as the dependent variable. Younger and older adult data were analyzed separately. Younger adult data is presented first followed by older adult data.

Overall, participants in the high expectancy condition ($M=58.00$, $SD=26.02$) depended more than participants in the low ($M=40.50$, $SD=22.85$) and standard ($M=35.25$, $SD=23.84$) expectancy conditions. The main effect of expectancy was statistically significant ($F(2, 54) = 4.99$, $p < .05$). Follow-up analyses revealed that the difference between the high and low expectancy condition was statistically significant ($t(19) = 17.50$, $p < .05$). Also, the difference between the high and standard expectancy conditions was statistically significant ($t(19) = 22.75$,

$p < .05$). There was no statistical difference between the low and standard expectancy conditions ($t(19) = 5.25, p < .05$). Figure 10 illustrates the main effect of expectancy. These results are consistent with the findings reported in Study 2 suggesting that, for younger adults, different expectations result in different levels of dependence on the automation.

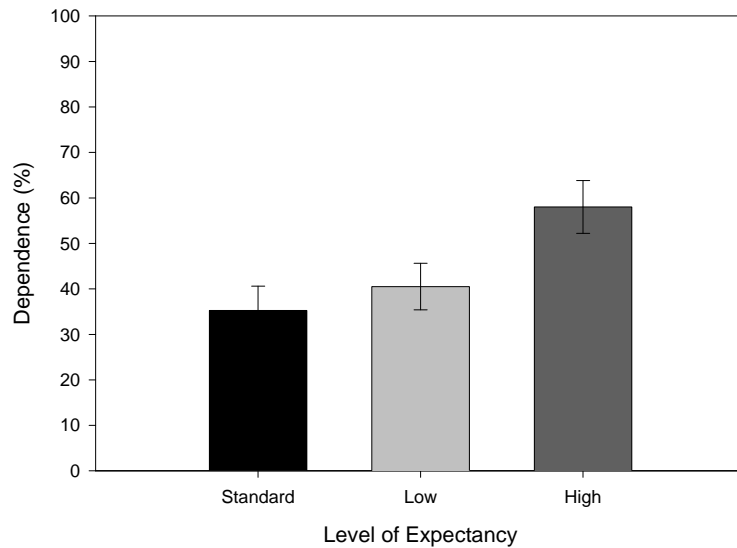


Figure 10: Younger adults' dependence during Block 1 by level of expectancy (low, high, and standard).

Younger adult data from Block 1 were plotted to display dependence over time (Figure 11). As was the case in Study 2, the pattern of data suggests that participants in the high, low, and standard conditions maintain relatively stable levels of dependence over time. Again, as in Study 2, it appears as though expectancy effects persist despite both automation errors and successes.

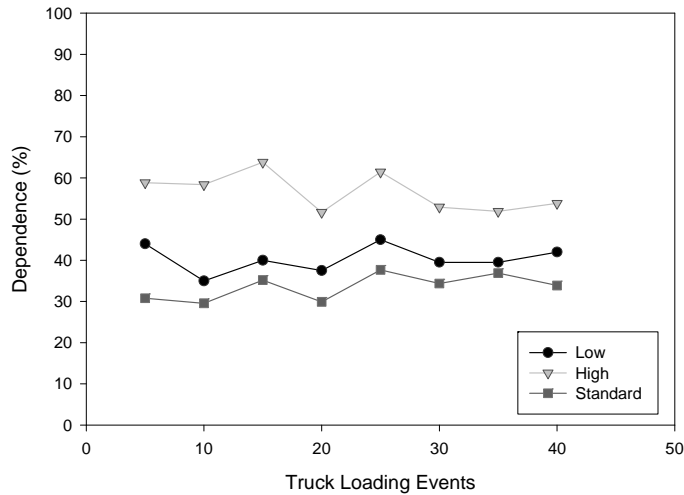


Figure 11: Younger adults' dependence over the course of Block 1 by level of Expectancy (low, high, standard).

Similarly to Study 2, there was no significant main effect of error type on dependence ($F(1, 54) = 2.51, p = .12$) although numerically, participants in the false alarm condition ($M=39.71, SD=27.39$) depended on the automation less than participants in the miss condition ($M=49.46, SD=23.54$). There was also no significant expectancy by error type interaction ($F(2, 54) = 1.29, p = .28$).

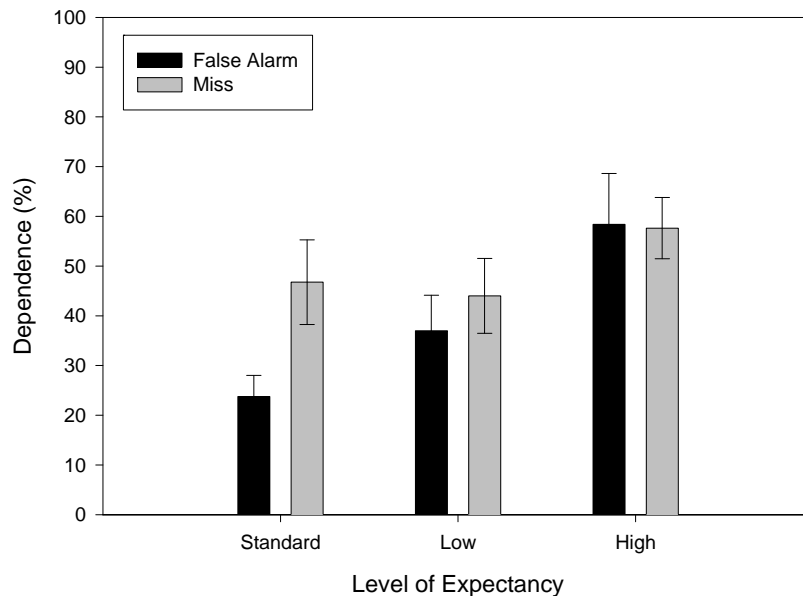


Figure 12: Younger adults' dependence during block 1 by expectancy and error type.

Older adult data were analyzed in the same way to determine if the results from the younger adult data generalize to older adult populations. Overall, there was a significant main effect of expectancy ($F(2, 54) = 3.22, p < .05$) (See Figure 13). Follow up tests revealed that participants in the high expectancy condition depended on the automation significantly more than participants in the standard expectancy condition ($t(19) = 21.38, p < .05$). However, there was no statistically significant differences between the high and low condition ($t(19) = 4.75, p = .59$) or between the low and standard conditions ($t(19) = 16.62, p = .07$). This pattern of data is different from the younger adult data.

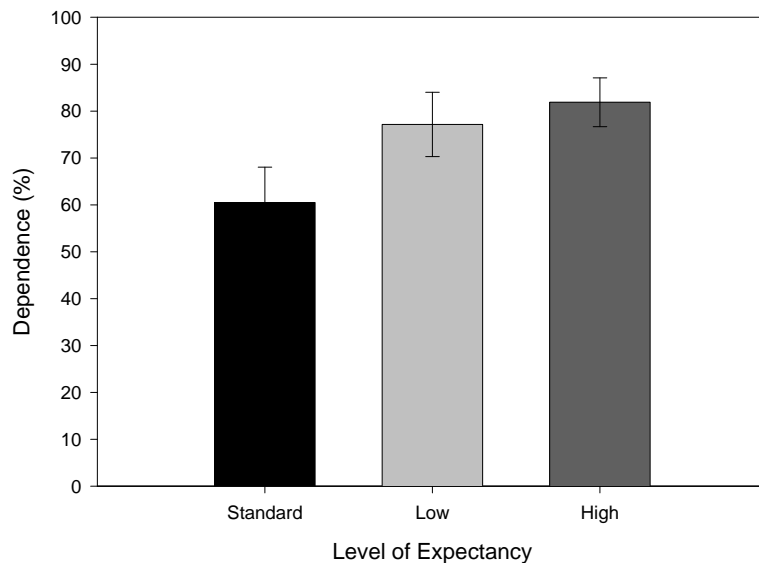


Figure 13: Older adults' dependence during Block 1 by level of expectancy.

Plotting older adults' dependence data over time shows that, similar to younger adults, the pattern of older adults' dependence remains relatively stable across Block 1 (Figure 14). However, at the end of Block 1, dependence behavior by participants in the low and high expectancy conditions begin to converge suggesting that expectancy effects in older adults may not be as strong as expectancy effects in younger adults.

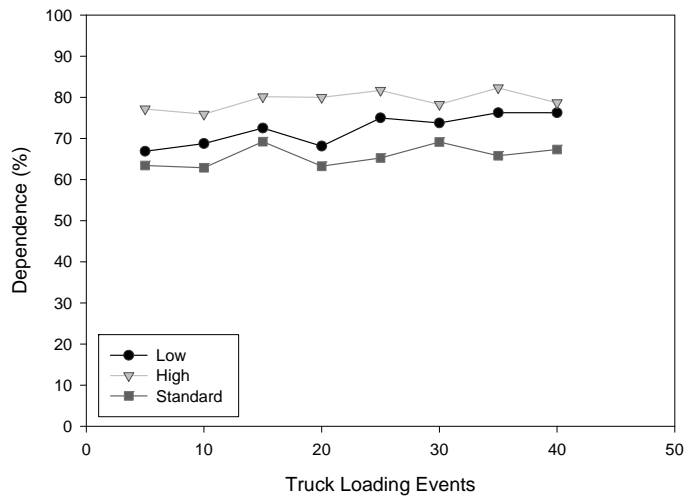


Figure 14: Older adults' dependence over the course of block 1 by level of expectancy.

There was also a significant main effect of error type for the older adult data ($F(1, 54) = 6.92, p < .05$). Overall, older adults in the false alarm condition ($M=63.67, SD=30.78$) depended on the automation less than older adults in the miss condition ($M=82.67, SD=27.41$). There was no significant expectancy by error type interaction ($F(2, 54) = 1.27, p = .29$).

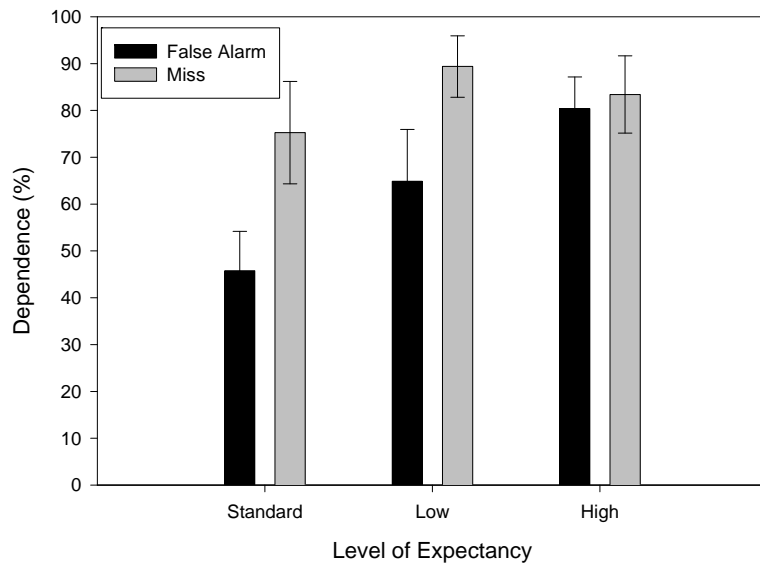


Figure 15: Older adults' dependence during Block 1 by level of expectancy and error type.

Reliance and Compliance during Block 1

Dependence data are a good measure of automation use, however, recent investigations have pointed out the importance of considering behavior during the alarm state (compliance) and the non-alarm state (reliance) of the automation separately (Dixon, Wickens, & McCarley, 2007). This approach is particularly useful when studying the effect of different automation errors. As above, younger and older adult data were analyzed separately and presented here in that order.

Compliance. Younger adult compliance data were analyzed using a factorial ANOVA with Expectancy and Error Type as between subject variables. Overall, participants in the High Expectancy condition ($M=48.13$, $SD=39.46$) complied with the automation more than participants in the Low ($M=39.00$, $SD=32.21$) and Standard ($M=27.38$, $SD=31.69$) expectancy conditions. There was a significant main effect of Expectancy ($F(2, 54) = 3.77$, $p < .05$). Participants in the High Expectancy condition complied significantly more than participants in the Standard Expectancy condition ($t(19) = 21.00$, $p < .05$). There were also significant differences between the High and Low expectancy condition, $t(19) = 15.75$, $p = .05$, however, there were no significant differences between the Low and Standard Expectancy conditions, $t(19) = 5.25$, $p = .51$.

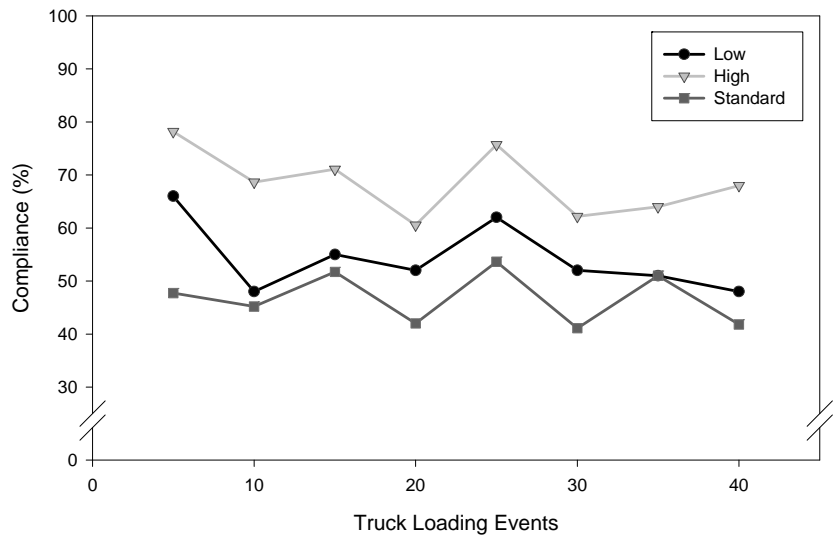


Figure 16. Younger adults' Block 1 compliance behavior by Level Expectancy (Low, High, and Standard).

Overall, younger adult participants in the False Alarm condition ($M=47.17$, $SD=28.05$) complied less than participants in the miss condition ($M=68.33$, $SD=24.80$), which was statistically significant ($F(1, 54) = 10.61$, $p < .05$). There was no significant interaction between Level of Expectancy and Error Type ($F(2, 54) = 1.31$, $p = .28$).

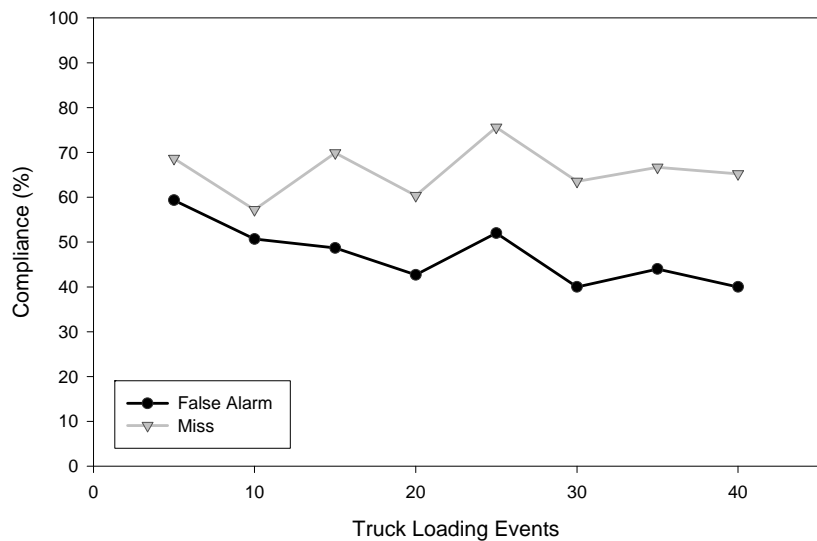


Figure 17. Younger adults' Block 1 compliance behavior by Error Type (False Alarm and Miss).

Older adult compliance data were analyzed in the same way as for younger adults.

Overall, participants in the Standard Expectancy condition ($M=62.25$, $SD=36.74$) complied with the automation less than participants in the High ($M=81.75$, $SD=30.81$) and Low ($M=81.13$, $SD=30.81$) Expectancy conditions, which was not statistically significant ($F(2, 54) = 2.53$, $p = .09$).

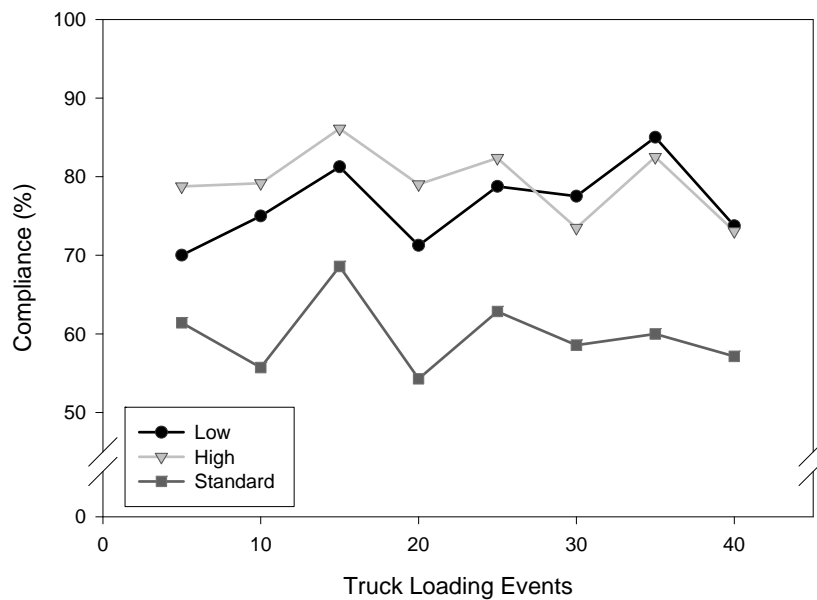


Figure 18. Older adults' Block 1 compliance behavior by Level of Expectancy (Low, High, and Standard).

Overall, older adults in the False Alarm condition ($M=63.33$, $SD=37.71$) complied with the automation less than older adults in the Miss condition ($M=86.75$, $SD=25.13$), which was statistically significant ($F(1, 54) = 8.48$, $p < .05$). There was no significant interaction between Expectancy and Error Type for older adults ($F(2, 54) = 1.16$, $p = .32$).

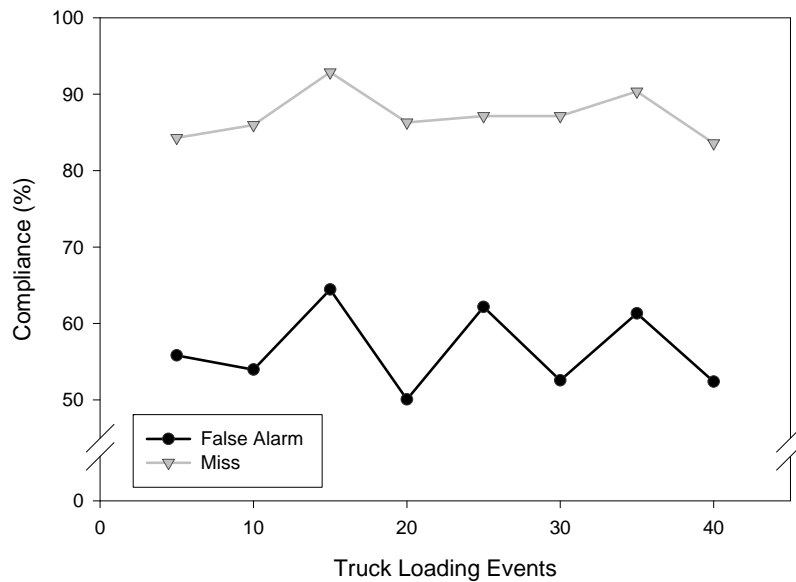


Figure 19. Older adults' Block 1 compliance behavior by Error Type (False Alarm and Miss).

Reliance. Younger adult reliance data were analyzed using a factorial ANOVA with Expectancy and Error Type as between subject variables. Overall, younger adults in the High expectancy condition ($M=46.00$, $SD=38.90$) relied more than younger adults in the Low ($M=26.38$, $SD=29.45$) and Standard ($M=21.50$, $SD=27.50$) expectancy conditions, which was statistically significant ($F(2, 54) = 3.20$, $p < .05$). Participants in the High expectancy condition relied significantly more than participants in the Standard expectancy condition ($t(19) = 24.50$, $p < .05$). There were no significant differences between the High and Low conditions ($t(19) = 19.63$, $p = .061$) or between the Low and Standard conditions ($t(19) = 4.88$, $p = .64$).

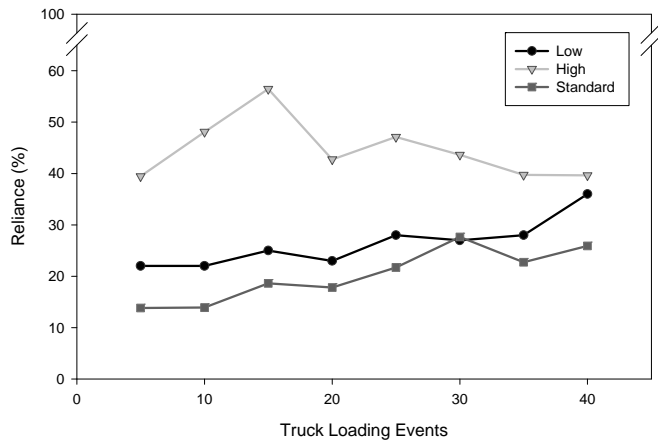


Figure 20. Younger adult Block 1 reliance by Level of Expectancy (Low, High, and Standard).

Overall, there was very little difference in reliance between young adults in the False Alarm condition ($M=32.00$, $SD=36.91$) and young adults in the Miss condition ($M=30.58$, $SD=30.40$), which was not statistically significant ($F(1, 54) = .03$, $p = .87$). The interaction was not statistically significant ($F(2, 54) = 1.34$, $p = .27$).

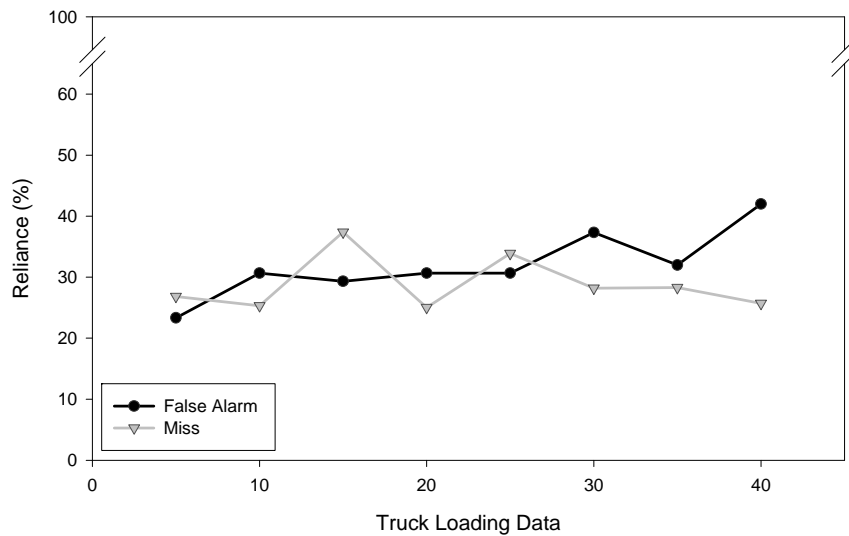


Figure 21. Younger adults' Block 1 reliance behavior by Error Type (False Alarm and Miss).

Reliance data for older adults in Block 1 revealed that there were no significant differences resulting from Level of Expectancy ($F(2, 54) = 2.29$, $p = .111$) or automation Error Type ($F(1, 54) = 2.43$, $p = .12$). The Expectancy by Error Type interaction was not statistically

significant ($F(2, 54) = .97, p = .38$).

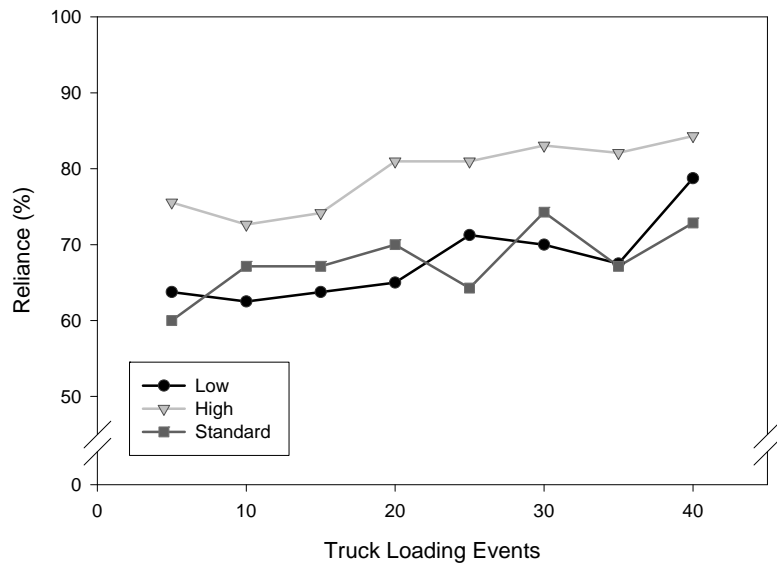


Figure 22. Older adults' Block 1 reliance behavior by Expectancy (Low, High, and Standard).

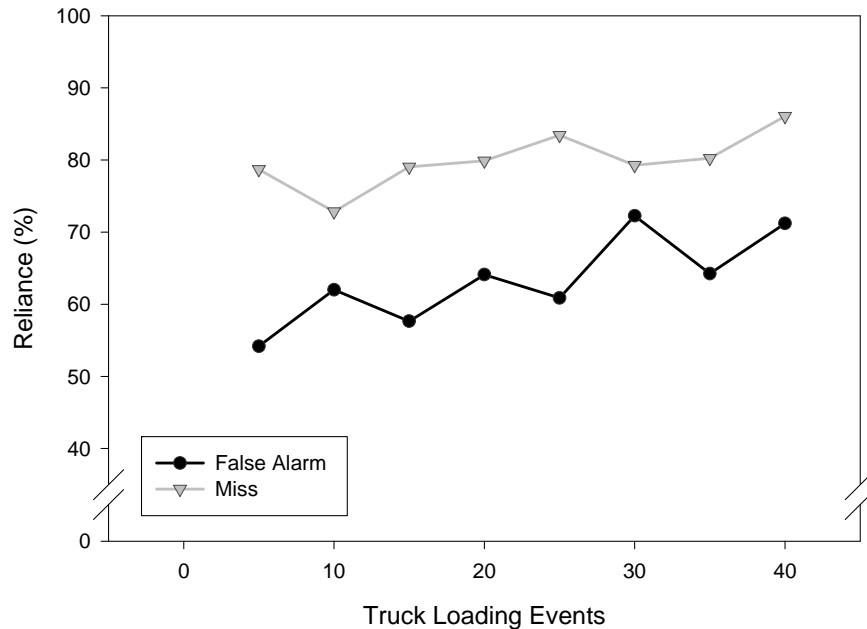


Figure 23. Older adults' Block 1 reliance behavior by Error Type (False Alarm and Miss).

Summary of Block 1. These results indicate that expectancy effects affect younger adults and that older adults are less susceptible to expectancy effects. The effect of error type was

significant for younger and older adults but only during the alarm state of the automation as evidenced by the compliance data.

Reliance and Compliance during Blocks 2 & 3

Data for Blocks 2 and 3 were analyzed only at the level of reliance and compliance since it is more descriptive to look at behavior during both the alarm and non-alarm states. Younger and older adult data were analyzed the same way as in Block 1.

Compliance. Overall, during Block 2, younger adults in the High expectancy condition (M=62.13, SD=33.59) complied more than younger adults in the Low (M=46.63, SD=30.00) and Standard (M=54.63, SD=30.13) expectancy conditions, which was not statistically significant ($F(2, 54) = 1.56, p = .22$). During Block 3, there was very little difference in compliance between the High (M=58.00, SD=37.35), Low (M=47.50, SD=30.56), and Standard (54.25, SD=31.89) expectancy conditions, which was not statistically significant ($F(2, 54) = .70, p = .50$).

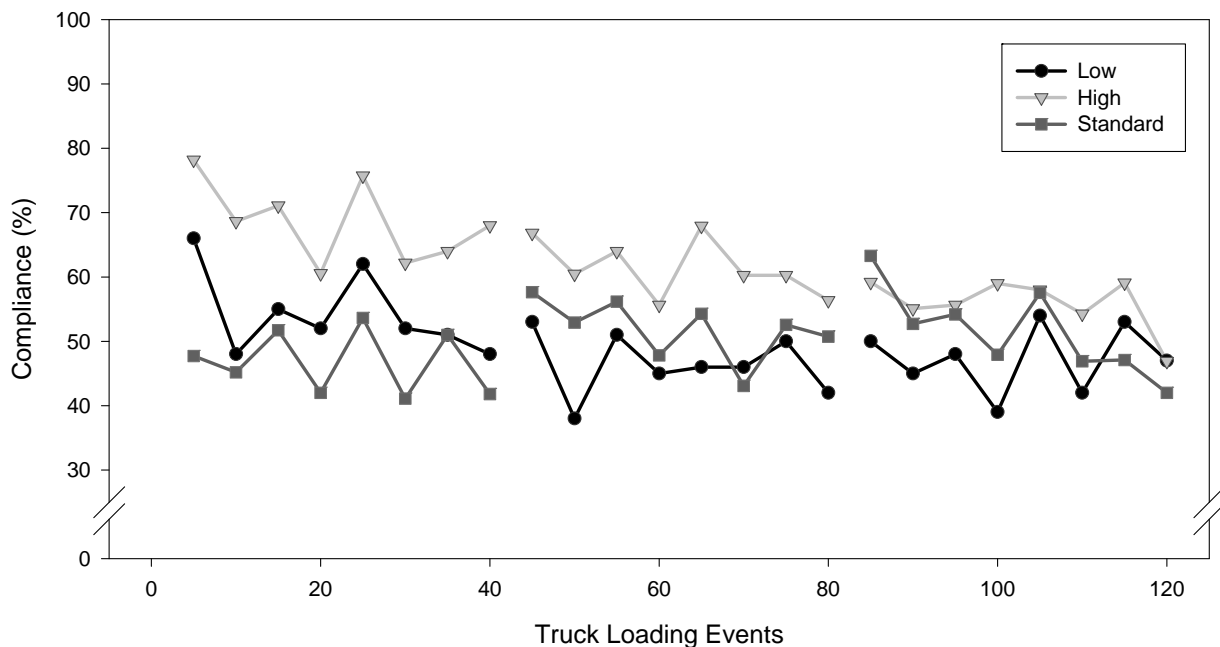


Figure 24. Younger adults' compliance behavior by Level of Expectancy (Low, High, and

Standard).

Overall, during Block 2, younger adults in the False Alarm condition ($M=39.17$, $SD=30.57$) complied with the automation less than participants in the Miss condition ($M=69.75$, $SD=24.26$), which was statistically significant ($F(1, 54) = 18.24, p < .05$). The interaction was not statistically significant ($F(2, 54) = .15, p = .86$). During Block 3, younger adults in the False Alarm condition ($M=35.00$, $SD=31.20$) complied less than participants in the Miss condition ($M=71.50$, $SD=23.86$), which was statistically significant ($F(1, 54) = 24.86, p < .05$). The interaction between Expectancy and Error Type was not significant in Block3 ($F(2, 54) = .118, p = .89$).

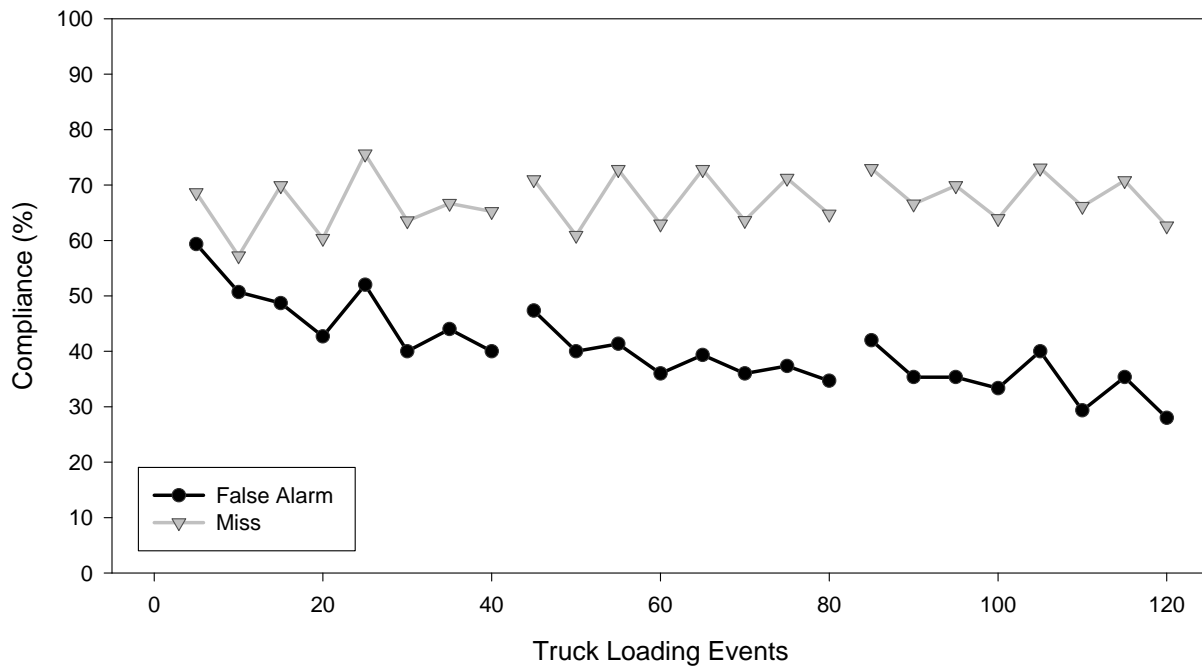


Figure 25. Younger adults' compliance behavior by Error Type (False Alarm and Miss).

Overall, during Block 2, older adults in the High ($M=77.13$, $SD=34.22$), Low ($M=80.13$, $SD=36.16$), and Standard expectancy conditions ($M=60.25$, $SD=38.28$) complied at a statistically equivalent level ($F(2, 54) = 2.20, p = .12$). Similar patterns were found in Block 3 for High ($M=73.25$, $SD=37.10$), Low ($M=77.38$, $SD=37.97$), and Standard ($M=57.5$, $SD=40.25$)

Expectancy conditions ($F(2, 54) = 2.01, p = .14$).

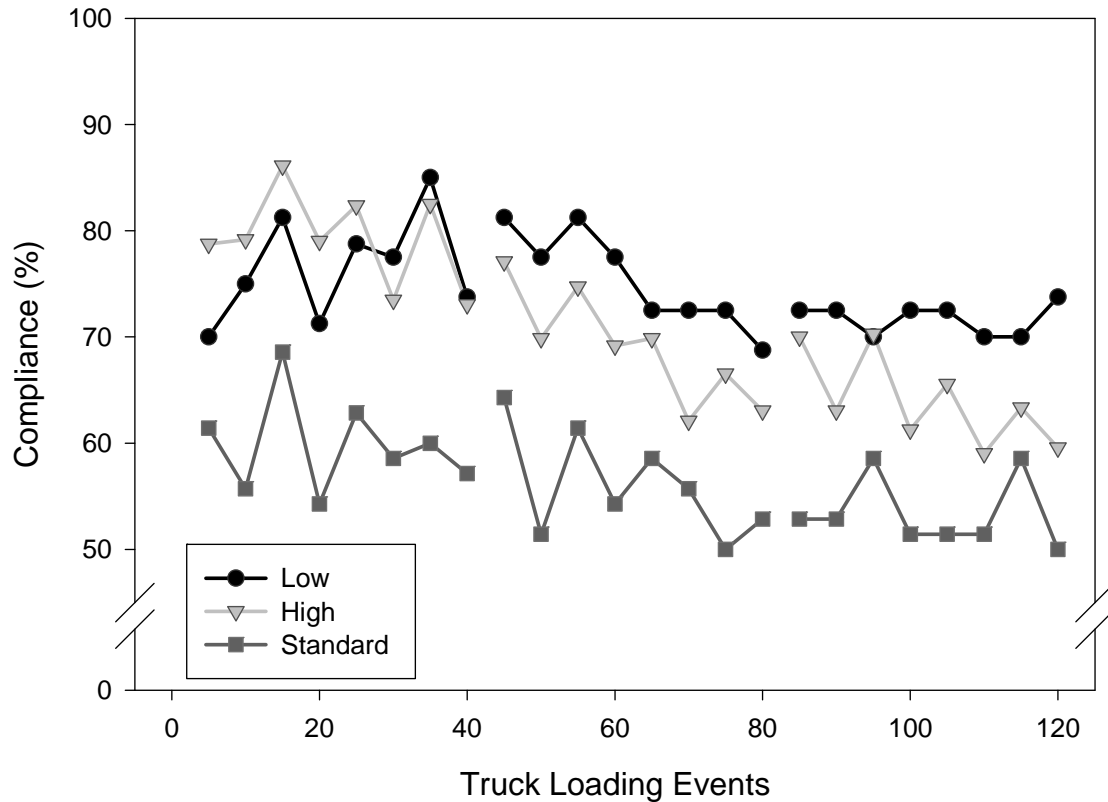


Figure 26. Older adults' compliance behavior by Level of Expectancy (Low, High, and Standard).

Older adults in the False Alarm condition ($M=56.42, SD=40.98$) complied less than participants in the Miss condition ($M=88.58, SD=22.99$) during Block 2 ($F(1, 54) = 14.86, p < .05$). During Block 3, older adults in the False Alarm condition ($M=50.08, SD=42.30$) complied less than older adults in the Miss condition ($M=88.67, SD=22.37$), ($F(1, 54) = 20.41, p < .05$). The interaction effect was not significant in Block 2 ($F(2, 54) = 1.45, p = .244$) or Block 3 ($F(2, 54) = 1.33, p = .27$).

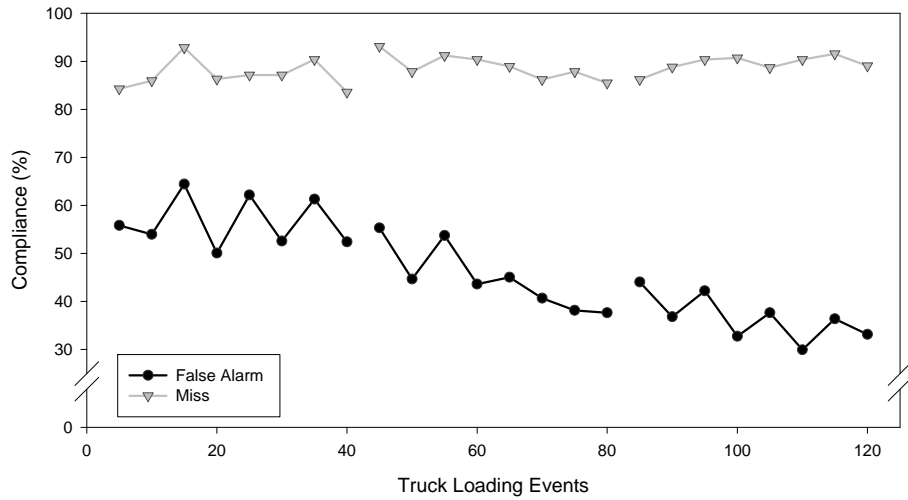


Figure 27. Older adults' compliance behavior by Error Type (False Alarm and Miss).

Reliance. Overall, there was no significant effect of Expectancy during Block 2 ($F(2, 54) = 1.82, p = .17$) or during Block 3 ($F(2, 54) = 2.43, p = .10$). In addition, there was no significant main effect of Error Type during Block 2 ($F(1, 54) = 2.13, p = .15$) or during Block 3 ($F(1, 54) = 2.43, p = .13$). Finally, there was no significant interaction during Block 2 ($F(2, 54) = .69, p = .50$) or during Block 3 ($F(2, 54) = 1.20, p = .31$).

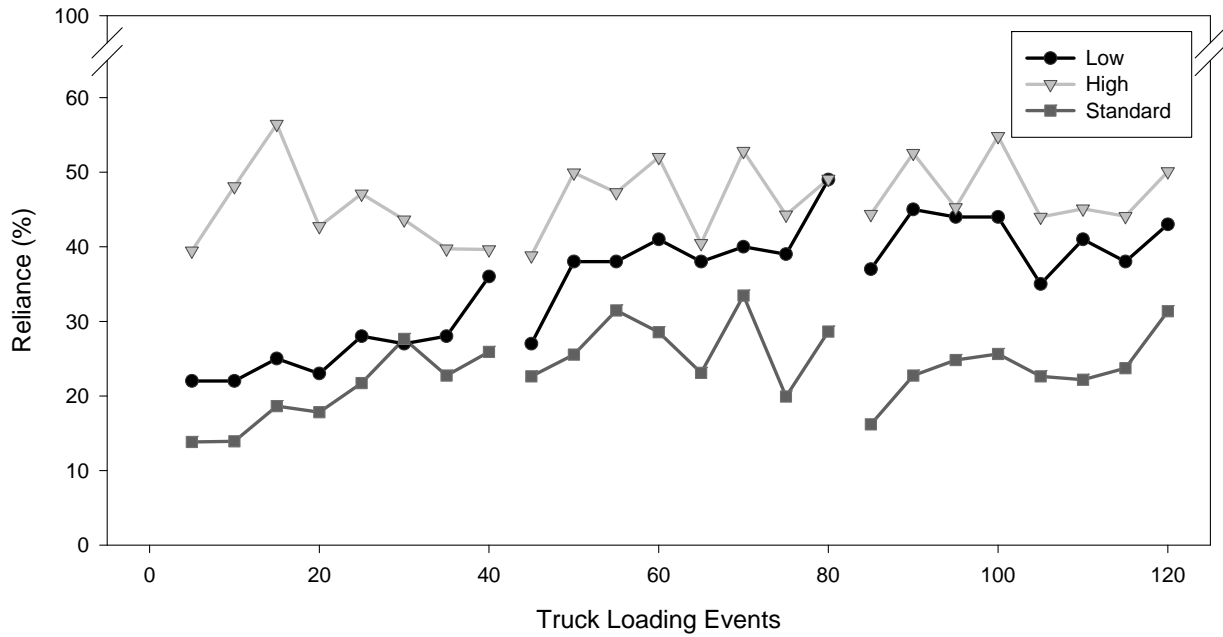


Figure 28. Younger adults' reliance behavior by Level of Expectancy (Low, High, and Standard).

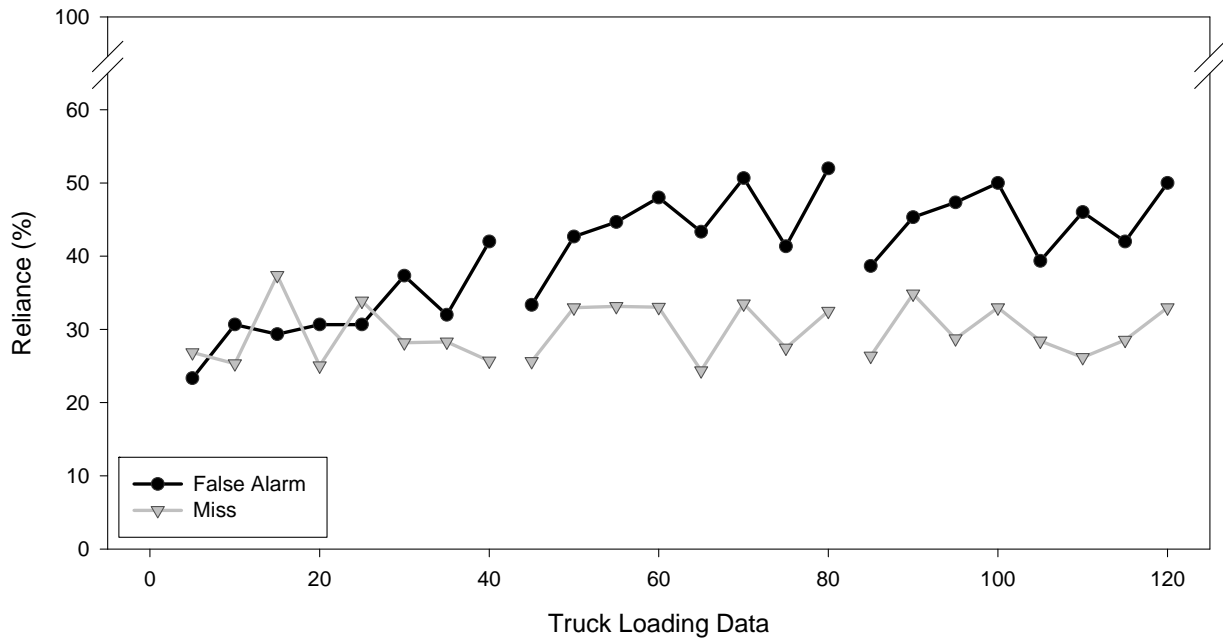


Figure 29. Younger adults' reliance behavior by Error Type (False Alarm and Miss).

Reliance data for older adults showed similar results to those found for younger adults.

There was no significant effect of Expectancy during Block 2 ($F(2, 54) = 2.44, p = .10$) or during

Block 3 ($F(2, 54) = 2.17, p = .12$). In addition, there was no significant main effect of Error Type during Block 2 ($F(1, 54) = .26, p = .62$) or during Block 3 ($F(1, 54) = .15, p = .70$). Finally, there was no significant interaction during Block 2 ($F(2, 54) = 1.21, p = .31$) or during Block 3 ($F(2, 54) = 1.23, p = .30$).

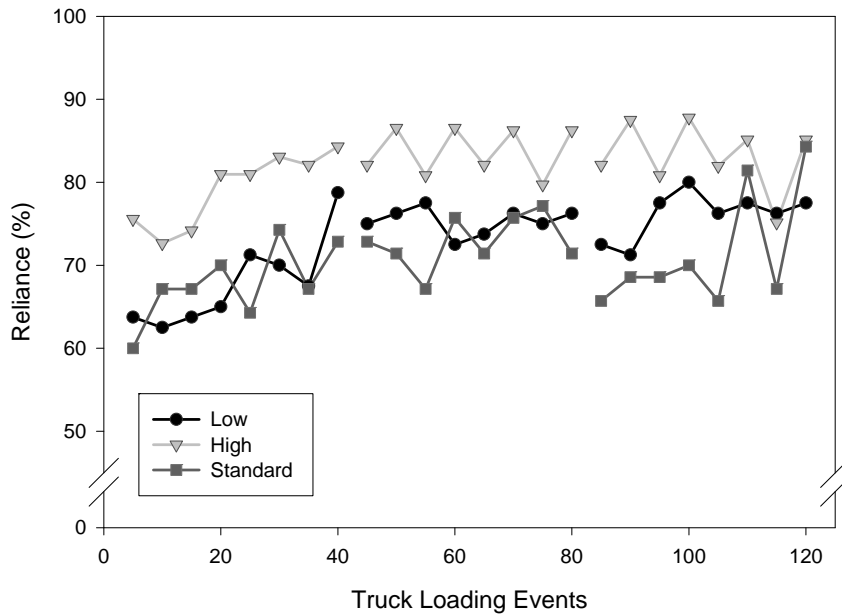


Figure 30. Older adults' reliance behavior by Level of Expectancy (Low, High, and Standard).

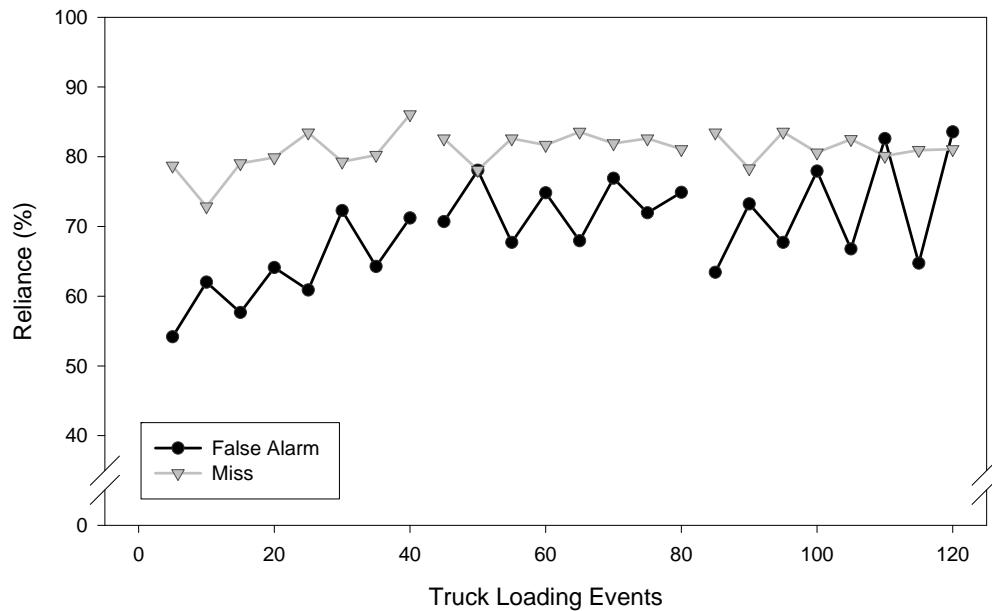


Figure 31. Older adults' reliance behavior by Error Type (False Alarm and Miss).

Summary of Blocks 2 & 3. Expectancy effects were not present during Block 2 or during Block 3 for either younger or older adults during the alarm state and the non-alarm state. In contrast, the compliance data for younger and older adults revealed a strong effect of Error Type during Blocks 2 and 3. The effect of Error Type was stronger for Block 3 compared to Block 2. In fact, the effect of Error Type was stronger in Block 2 compared to Block 1 for the compliance data suggesting some knowledge being accrued regarding Error Type during the alarm state of the automation.

Reliance and Compliance during Block 4

Block 4 was different from the first three blocks because the reliability of the automation was only 60% compared to 90% during the first three blocks. The purpose of Block 4 was to help determine sources of behavior changes that may have emerged during the first three blocks. The data were analyzed in the same fashion as the data from the first three blocks.

Compliance. Overall, there was no effect of expectancy for younger adults during Block

4 ($F(2, 54) = 1.87, p = .16$). There was a strong Error Type effect ($F(1, 54) = 122.02, p < .05$).

Overall, young adults in the False Alarm condition ($M=33.92, SD=17.33$) complied with the automation less than participants in the Miss condition ($M=78.58, SD=13.95$). There was no significant interaction effect ($F(2, 54) = .40, p = .68$).

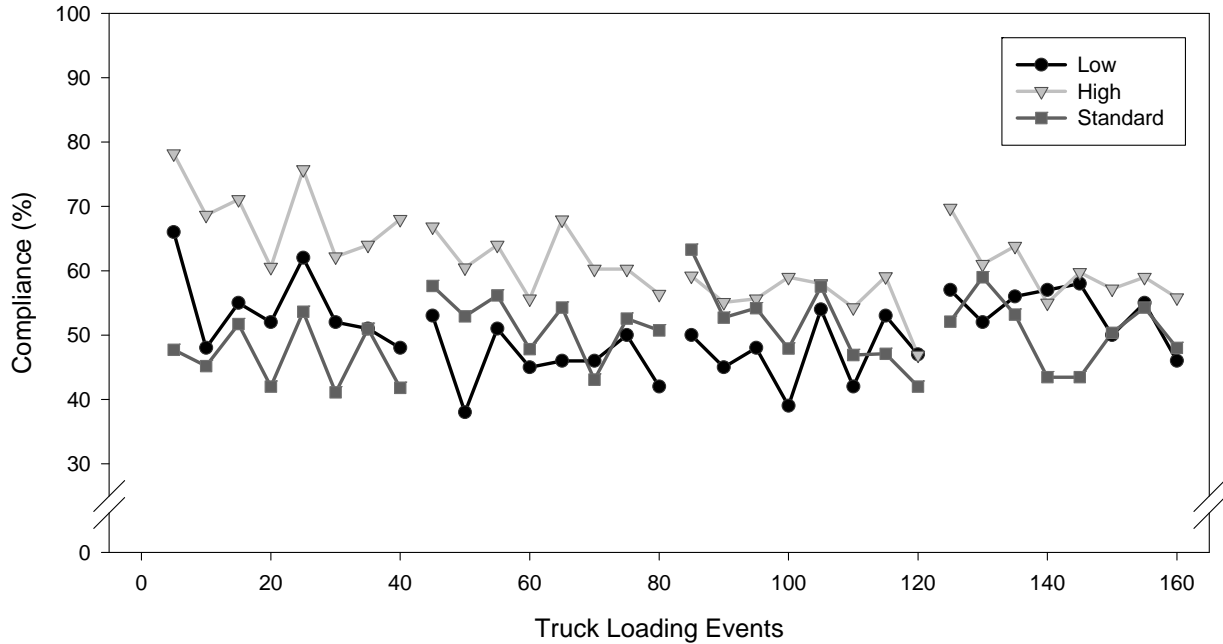


Figure 32. Younger adults' compliance behavior by Level of Expectancy (Low, High, and Standard).

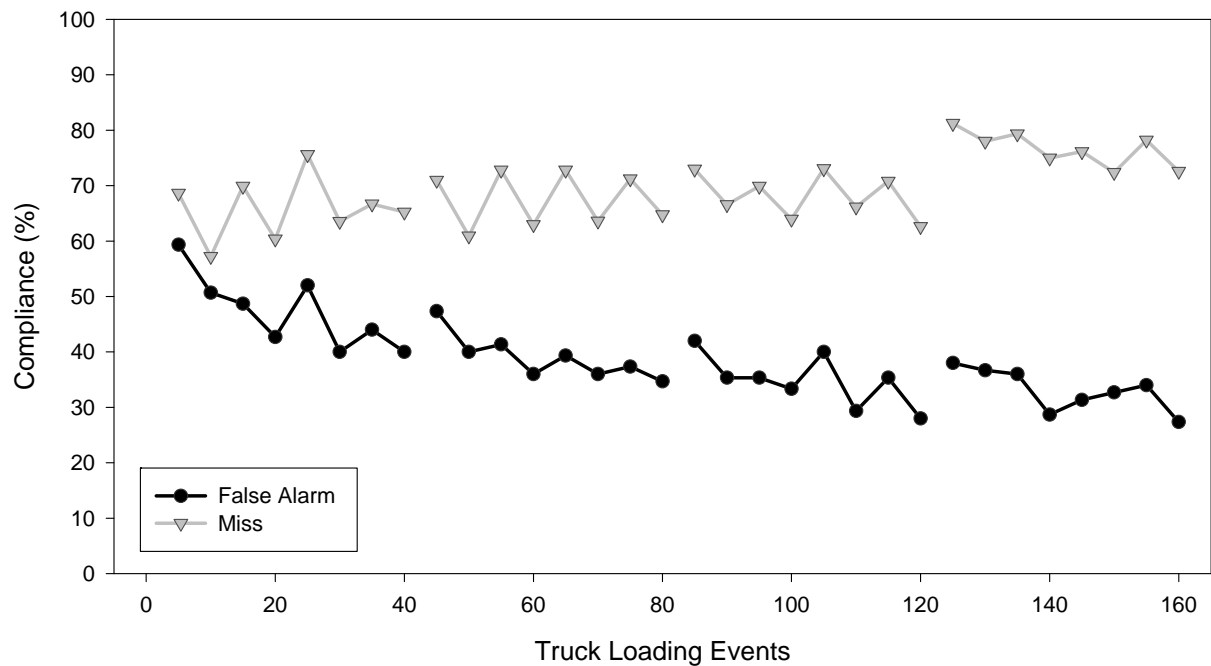


Figure 33. Younger adults' compliance behavior by Error Type (False Alarm and Miss).

There was no effect of expectancy for older adults during Block 4 ($F(2, 54) = 1.12, p = .34$). There was a strong Error Type effect ($F(1, 54) = 26.13, p < .05$). Overall, older adults in the False Alarm condition ($M=54.75, SD=38.88$) complied with the automation more than older adults in the Miss condition ($M=93.00, SD=13.78$). There was no significant interaction effect ($F(2, 54) = 1.25, p = .30$).

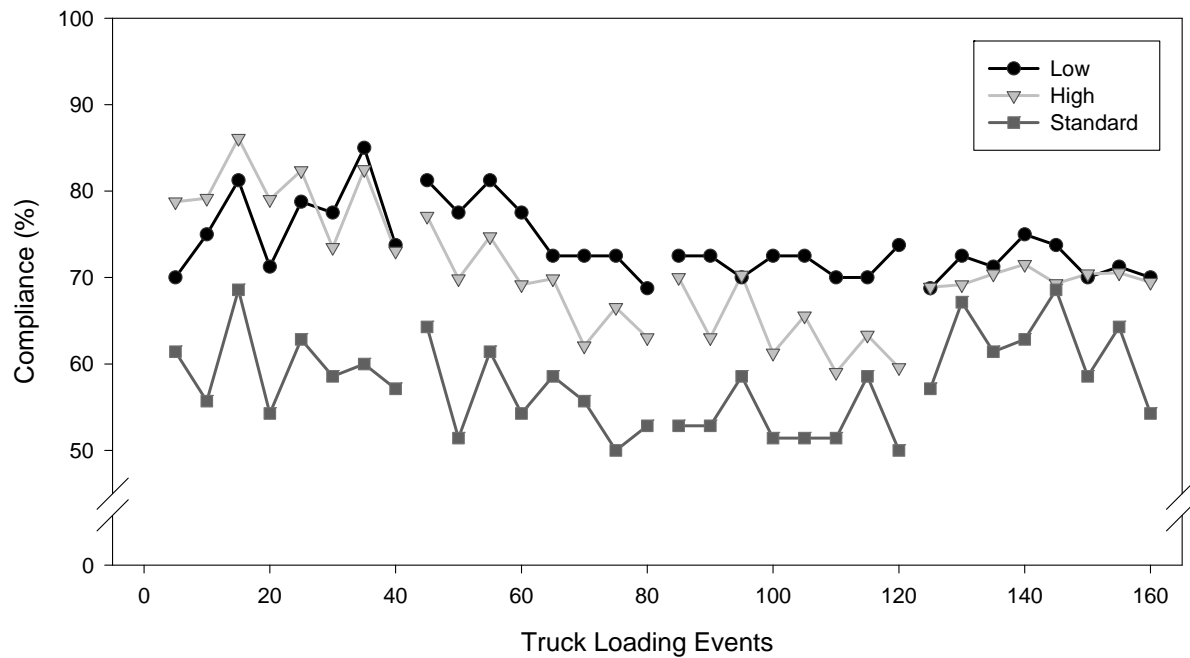


Figure 34. Older adults' compliance behavior by Level of Expectancy (Low, High, and Standard).

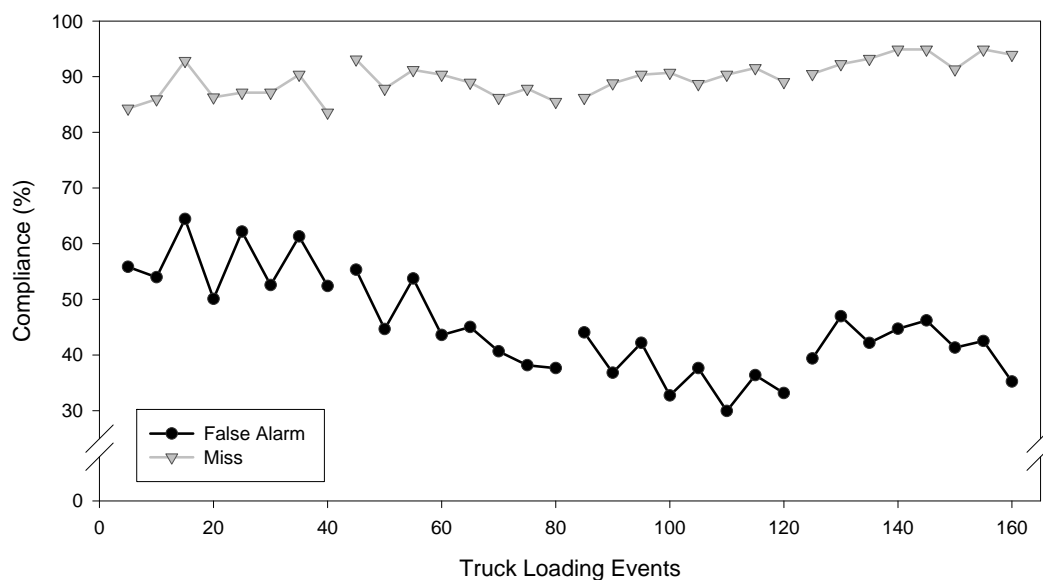


Figure 35. Older adults' compliance behavior by Error Type (False Alarm and Miss).

Reliance. There were no significant effects of Expectancy ($F(2, 54) = .986, p = .38$) or Error Type ($F(1, 54) = 1.50, p = .23$) for younger adults. The Expectancy by Error Type

interaction was not statistically significant ($F(2, 54) = 1.69, p = .19$). Similar patterns of data were found for older adults. There were no significant effects of Expectancy ($F(2, 54) = 2.42, p = .10$) or Error Type ($F(1, 54) = 1.55, p = .22$). The Expectancy by Error Type interaction was not statistically significant ($F(2, 54) = 1.06, p = .35$).

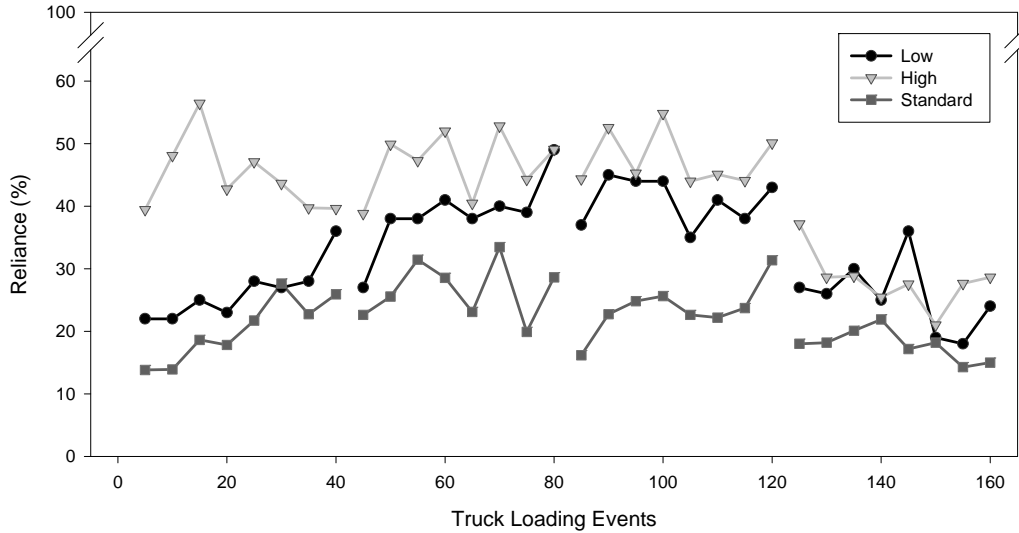


Figure 36. Younger adults' reliance behavior by Level of Expectancy (Low, High, and Standard).

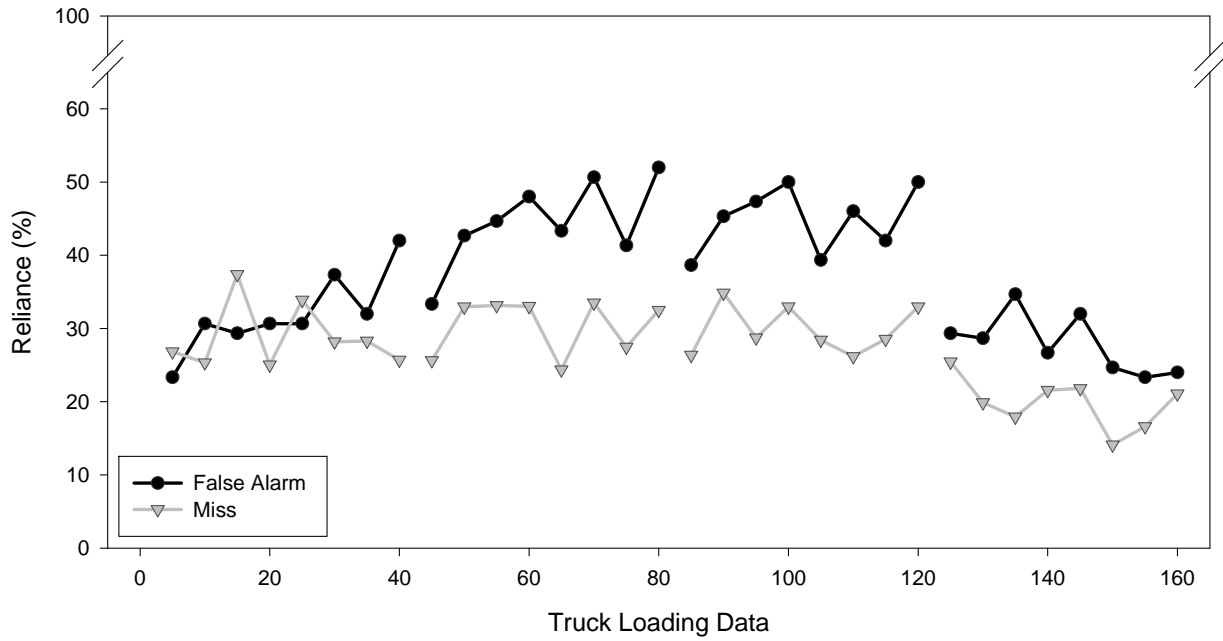


Figure 37. Younger adults' reliance behavior by Error Type (False Alarm or Miss).

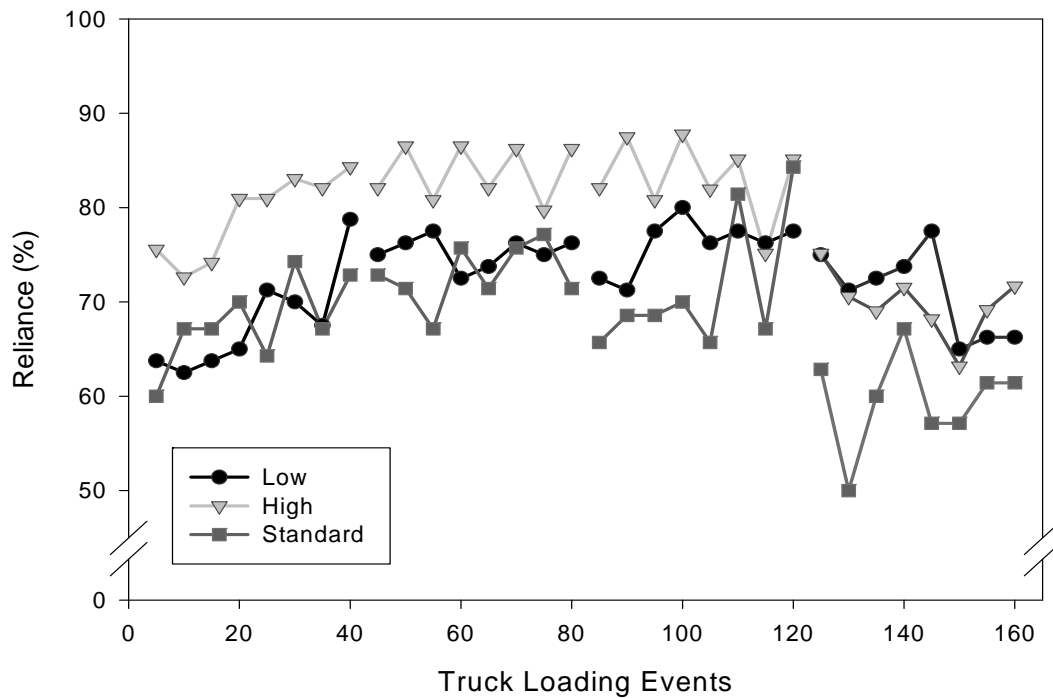


Figure 38. Older adults' reliance behavior by Level of Expectancy (Low, High, and Standard).

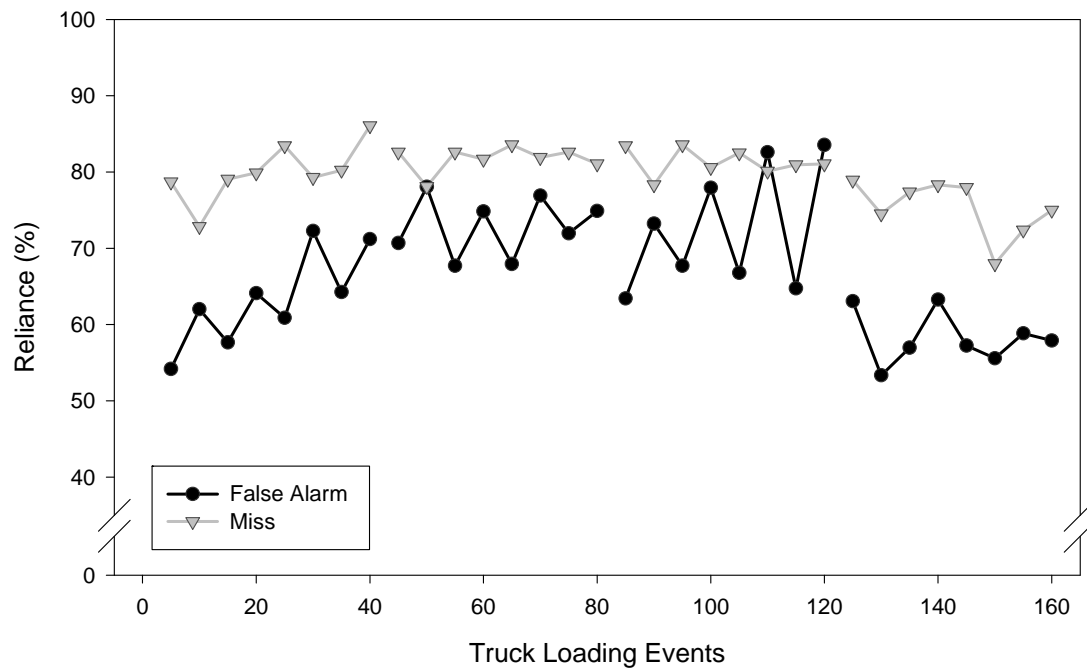


Figure 39. Older adults' reliance behavior by Error Type (False Alarm and Miss).

Summary of Block 4 Data. The purpose of Block 4 was to try to better understand how behavior is guided when interacting with automation. By transferring participants to a less reliable version of the automation, we would be able to assess changes in behavior from the previous blocks that might shed light on the nature of reliance and compliance. Overall, we found that expectancy effects were non-existent for younger and older adults regardless of whether we measured behavior during the alarm or non-alarm state. There were very strong Error Type effects when we measured behavior during the alarm state. In contrast, there were no Error Type effects when we measured behavior during the non-alarm state of the automation. Taken together, these data suggest that Error Type differences provide important information to participants to help guide behavior but the Error Type information only informs participants in the context of an automation alert.

DISCUSSION

In Study 2, findings suggested that high and low expectancies had a strong effect on dependence in a younger adult population, increasing and reducing dependence respectively. However, that preliminary study used only a 30-minute experimental block. The goals for the current study were to see if (a) the expectancy effect could be replicated and if (b) the expectancy effect extended beyond 30 minutes. In addition, age-related effects were investigated since older adults stand to benefit greatly from automation and age-related investigations of expectancy effects are missing in the literature.

The results show that the current study successfully replicated Study 2. Younger adults in the high expectancy condition, during the first block, depended on the automation more than participants in the low and standard conditions. These differences were achieved using a fairly subtle manipulation of the automation description. These results differ from Madhavan and Wiegmann (2007) who did not find any expectancy effects for 90 percent reliable systems. One explanation for the difference is that the Madhavan and Wiegmann task was a binary detection luggage screening task where participants had to simply decide to stop or pass a piece of luggage. The relatively simple nature of the task may have allowed participants more attention for recognizing the high reliability of the automation. In contrast, the dual-task used in the current study was high paced and attention-demanding which may have prevented participants from accruing sufficient knowledge of system reliability requiring them to rely on their expectations to guide behavior.

The effect of expectancy on dependence had dissipated by Block 2 suggesting that expectancy effects may be short lived and fail to extend beyond a 20-30 minute time period. Madhavan and Wiegmann (2007) found a similar result for a 70 percent reliable system and

attributed it to a violation of the “perfect automation schema” originally proposed by Dzindolet et al. (2001). When participants interacted with an expert automated system, they dramatically reduced their reliance and compliance in the face of automation errors (Madhavan and Wiegmann). This may indeed be the case except that in the current study, participants’ expectations were manipulated to shift participants in the low expectancy condition away from a “perfect automation schema.” It should be stressed that a short-lived effect of expectancy should not be equated with an unimportant effect of expectancy. In certain domains even a single error can be catastrophic.

Older adults did not exhibit statistically significant expectancy effects. However, numerically, there were differences in dependence that were similar to younger adults. Overall, there was a great deal of variance in the data particularly with the older adults. The likely source of the variance was the vague understanding participants had regarding the functioning of the automation. As mentioned previously, most automation studies provide participants with knowledge of system functioning in the form of either the type of errors the automation commits, the overall reliability of the system, or a combination of the two, which plausibly leads to less variability. The current study was specifically designed to ecologically sample the uncertainty people have when interacting with unfamiliar systems, so the system descriptions were purposely vague. The potential downside of this approach, however, is that people may develop an understanding of system performance in different ways and at different rates which results in higher variance.

One advantage of providing participants with explicit information related to system errors and system reliability is that it can reduce variance. However, are participants being biased to depend less on the automation because of this information? Results from Study 3 suggest that

yes, providing participants with some knowledge of system performance does degrade their overall dependence. Both younger and older adults in the standard expectancy condition depended, relied, and complied with the automation at a lower rate than participants in the high and low expectancy conditions, despite reporting equivalent expectancies to the high expectancy condition on the expectancy questionnaire. One explanation is that because participants in the standard condition had some understanding of system errors, they were better able to recognize and encode automation errors, resulting in an underestimation of automation reliability.

There has been debate in the literature regarding whether system misses or false alarms are more detrimental to human-automation interaction (e.g., Johnson, 2004; Meyer, 2004; Dixon, Wickens, & McCarley, 2007). Interestingly, results from Study 3 suggest that perhaps it is not the type of error per se that has a greater impact on behavior, but rather the state the automation is in when the error occurs. During the alarm state, when the alarm was always correct (which was the case for participants in the miss condition), participants complied with the automation significantly more than when the alarm was not always correct (which was the case for participants in the false alarm condition). Conversely, during the non-alarm state, when the automation was functioning perfectly (which was the case for participants in the false alarm condition), there was no reliance difference from when the automation made errors (which was the case for participants in the miss condition). If it was the error type that was influencing behavior, we would expect differences at both the reliance AND compliance levels of analysis; we only found differences at the compliance level.

Taking the results from Study 3 as a whole, the patterns of data suggest that younger adults are more susceptible to expectancy effects compared to older adults. Expectancy effects seem to be robust to errors (as the effect was sustained across the first block) but not robust to

time and experience (as the effect disappeared by second block). Participants gained some knowledge of system functioning evidenced by the increasing Error Type affect found across blocks. But, as stated previously, this knowledge did not seem to extend beyond the alarm state of the automation.

Chapter 6 – General Discussion

The three studies presented in this report were designed to understand the precursor variables that influence behavioral acceptance of technology. We identified a gap in the automation literature: little was known about the extent to which user expectations affect use of an automated system. Study 1 investigated the extent to which subtle descriptions of automation can change the reported expectations of that system's likely performance. Study 2 was a preliminary investigation of how expectations, once manipulated, affect general dependence on automation. In addition, Study 2 began to investigate differential effects of automation false alarms and misses on dependence. Study 3 expanded on Study 2 by extending the experimental session, recording more specific measures of automation use (reliance and compliance), using a more dynamic and attention-demanding task, and including older adults.

Taken as a whole, these studies make five important contributions. (1) Expectations can be easily manipulated and people are susceptible to subtle differences in automation descriptions. (2) Expectations do result in quantitatively different dependence behavior. (3) Expectancy effects can withstand some failures and successes by the automation but over time interacting with the automation expectancy effects dissipate. (4) With experience, participants draw on other sources of information to help guide behavior; in this case, information regarding errors during the alarm state of the automation guide behavior. (5) Older and younger adults show very similar patterns of behavior, although older adults tended to comply and rely on the automation to a greater degree compared to younger adults and younger adults appear more sensitive to expectancy manipulations.

The findings from these three studies have important implications for Deere & Company especially with respect to issues of technology acceptance. Understanding the ease with which

expectations can be changed through system descriptions will have impacts on both training of operators and marketing of products. Recognizing that more system knowledge is accrued by operators when the state of the automation is salient will have important design implications. Also, understanding that presenting new users with specific knowledge of potential errors can inappropriately decrease dependence on the system should influence how operators are trained on Deere's automated technology products.

References

- Bowers, C. A., Oser, R. L., Salas, E., & Cannon-Bowers, J. A. (1996). Team performance in automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications* (pp. 163-181). Mahwah, NJ: Lawrence Erlbaum.
- Cantor, N., & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology*, *35*, 38-48.
- Chappell, S. L. (1997). The effects of experience and automation on failure detection. *Proceedings of the 9th International Symposium on Aviation Psychology*, 773-779.
- Clark, M. C., Czaja, S. J., & Weber, R. A. (1990). Older adults and daily living task profiles. *Human Factors*, *32*, 537-549.
- Cohen, M. S., Parasuraman, R., & Freeman, J. T. (1998). Trust in decision aids: A model and its training implications. *Proceedings, 1998 Command and Control Research and Technology Symposium*, 1-37. Department of Defense C4ISSR Cooperative Research Program.
- Craik, F. I. M., & Salthouse, T. A. (Eds.). (2000). *The Handbook of Aging and Cognition* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Dixon, S. R. & Wickens, C. D. (2006). Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence in high workload. *Human Factors*, *48*, 474-486.
- Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007) On the independence of compliance and reliance: Are automation false alarms worse than misses? *Human Factors*, *49*, 564-572.
- Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (1999). Misuse and disuse of automated aids. *Proceedings of the Human Factors and Ergonomics Society*, *43rd Annual*

Meeting, 339.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, 58, 697-718.

Feldman, R. S., & Theiss, A. J. (1982). The teacher and student as Pygmaliions: Joint effects of teacher and student expectations. *Journal of Educational Psychology*, 74, 217-223.

Fisk, A. D., & Rogers, W. A. (2002). Psychology and aging: Enhancing the lives of an aging population. *Current Directions in Psychological Science*, 11, 107-110.

Gottlob, L. R. (2006). Age-related deficits in guided search using cues. *Psychology and Aging*, 21, 526-534.

Harris, M. J., & Rosenthal, R. (1985). Mediation of interpersonal expectancy effects: 31 meta-analyses. *Psychological Bulletin*, 97, 363-386.

Hernandez-Peck, M. (2001). Older farmers: Factors affecting their health and safety.

Retrieved September 29, 2009 from the World Wide Web:

<http://www.nasdonline.org/document/1825/d001760/older-farmers-factors-affecting-their-health-and-safety.html>.

James, W. (1981). The principles of psychology. In F. H. Burkhardt, F. B. Bowers, & I. K. Skrupskelis (Eds.), *The works of William James* (Vol. I). Cambridge, MA: Harvard University Press. (Original work published in 1890)

Jamieson, D. W., Lydon, J. E., Stewart, G., & Zanna, M. P. (1987). Pygmalion revisited: New evidence for student expectancy effects in the classroom. *Journal of Educational Psychology*, 79, 461-466.

- Jian, J., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, 4, 53-71.
- Johnson, J. D. (2004). *Type of automation failure: The effects on trust and reliance in automation*. Unpublished master's thesis, Georgia Institute of Technology, Atlanta.
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46, 50-80.
- Lee, J. D. (2006). Human factors and ergonomics in automation design. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics: Third Edition* (pp. 1570-1596). Hoboken, NJ: John Wiley & Sons, Inc.
- Madhavan, P. & Wiegmann, D. A. (2007). Effects on information source, pedigree, and reliability on operator interaction with decision support systems. *Human Factors*, 49, 773-785.
- Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Human Factors*, 43, 563-572.
- Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human Factors*, 46, 196-204.
- Riley, V. (1996). Operator reliance on automation: Theory and data. In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications* (pp. 19-35). Mahwah, NJ: Lawrence Erlbaum.
- Rosenthal, R. (1966). *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts.
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart,

and Winston.

- Rothbart, M., Evans, M., & Fulero, S. (1979). Recall for confirming events: Memory processes and the maintenance of social stereotypes. *Journal of Experimental Social Psychology, 15*, 343-355.
- Salthouse, T. A. (1992). Working-memory mediation of adult age differences in integrative reasoning. *Memory & Cognition, 20*, 413-423.
- Sanchez, J. (2005) *Human-automation interaction: factors that affect human behavior and system performance*. Unpublished manuscript, Georgia Institute of Technology.
- Sanchez, J. (2006). *Factors that affect trust and reliance on an automated aid*. Unpublished doctoral dissertation, Georgia Institute of Technology, Atlanta.
- Sanchez, J., Ezer, N., Rogers, W. A., & Fisk, A. D. (2006). Estimating reliability of automated aids: Effects of age and system reliability changes. *Cognitive Technology, 11*, 5-13.
- Siedlecki, K. L., Salthouse, T. A., & Berish, D. E. (2005). Is there anything special about aging of source memory? *Psychology and Aging, 20*, 19-32.
- Shipley, W. C. (1986). *Shipley Institute of Living Scale*. Los Angeles: Western Psychological Services.
- Snyder, M. L., & Frankel, A. (1976). Observer bias: A stringent test of behavior engulfing the field. *Journal of Personality and Social Psychology, 34*, 857-864.
- Stephan, W. G. (1985). Intergroup relations. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology: Vol. 2*, (3rd ed., Vol. 2, pp. 599-658). United States of America: Newbery Award Records, Inc.
- Van Ittersum, K., Rogers, W. A., O'Brien, M. A., Caine, K. E., Parsons, L. J., & Fisk, A. D. (2009). *Understanding technology acceptance: Phase IV – Testing the predictive validity*

- (HFA-TR-0905). Atlanta, GA: Georgia Institute of Technology, School of Psychology, Human Factors and Aging Laboratory.
- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale III. (3rd Ed.)*. San Antonio, TX: The Psychological Corporation.
- Wickens, C. D., Helleberg, J., & Xu, X. (2002). Pilot maneuver choice and workload in free flight. *Human Factors, 44*, 171-188.
- Wickens, C. D., & Xu, X. (2002). *Automation trust, reliability and attention* (Tech. Rep. No. AHFD-02-14/MAAD-02-2). Savoy, Illinois: University of Illinois at Urbana-Champaign, Aviation Human Factors Division, Institute of Aviation.
- Wickens, C. D., & Carswell, C. M. (2006). Information processing. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics: Third Edition* (pp. 1570-1596). Hoboken, NJ: John Wiley & Sons, Inc.
- Wilkison, B. D, Fisk, A. D., & Rogers, W. A. (under review). 'Enough information to be dangerous': Mental model quality and collaborative system performance. *Ergonomics*.

Appendix A: Automation Attitudes

An automated system is a technologically-based system used to partially or fully assist the human in tasks involving sensing, detecting, information processing, making decisions and/or executing actions.

Please circle the number that best describes your feeling, impression, or belief.

1. Automated systems are deceptive

1 2 3 4 5 6 7

Not at All

Extremely

2. Automated systems behave in an underhanded manner

1 2 3 4 5 6 7

Not at All

Extremely

3. I am suspicious of automated systems' intent, action, or outputs

1 2 3 4 5 6 7

Not at All

Extremely

4. I am wary of automated systems

1 2 3 4 5 6 7

Not at All

Extremely

5. Automated systems' actions have a harmful or injurious outcome

1 2 3 4 5 6 7

Not at All

Extremely

6. I am confident in automated systems

1 2 3 4 5 6 7

Not at All

Extremely

7. Automated systems provide security

1 2 3 4 5 6 7

Not at All

Extremely

8. Automated systems have integrity

1 2 3 4 5 6 7

Not at All

Extremely

9. Automated systems are dependable

1 2 3 4 5 6 7

Not at All

Extremely

10. Automated systems are reliable

1 2 3 4 5 6 7

Not at All

Extremely

11. I can trust automated systems

1 2 3 4 5 6 7

Not at All

Extremely

12. I am familiar with automated systems

1 2 3 4 5 6 7

Not at All

Extremely

13. To what extent do you think you could count on an Automated System to do its job?

1 2 3 4 5 6 7

Not at All

Completely

14. Overall, how much would you trust an Automated System?

1 2 3 4 5 6 7

Not at All

Completely

15. Please indicate how often you think an Automated System would provide correct information (using a %).

(Example: I think an Automated System would provide correct information ##% of the time)

_____ %

Appendix B: Expectancy Descriptions

Low-Expectancy Description

An Automated Warehouse Management System is a system that scans the inside of truck trailers, calculates the amount of space available in the truck, loads shipments onto the truck, determines if the truck is full, and when the truck is full notifies the Warehouse Manager to dispatch the truck. The Automated Warehouse Management System that is being tested today is designed for large, high traffic commercial warehouse operations

SRT-1 Automated Warehouse Management System

We are working with a company on issues of automation, as well as being funded by the National Institute of Health for this work. Let me tell you a little about the system you will be helping us test. The company first became involved in sensory technologies in 2000 with the sole mission of creating advanced scanning and decision making systems for warehouse loading and shipping applications. In 2001, the company proposed an Automated Warehouse Management System called the SRT and in 2004 proposed a Smart Automated Warehouse Management System, the SRT -1. The company's first prototype system, the SRT-1, utilizes advanced decision algorithms and sensing technologies that have the ability to adjust to differing warehouse and loading conditions. Testing of the SRT-1 has not begun so designers are unsure of the accuracy, reliability, and robustness of the Automated Warehouse Shipping system and how it will compare to the industry standard. Because this is a first prototype Automated Warehouse Management System, it is expected that the SRT-1 will perform at a low level with some performance errors.

High Expectancy Condition

An Automated Warehouse Management System is a system that scans the inside of truck trailers, calculates the amount of space available in the truck, loads shipments onto the truck, determines if the truck is full, and when the truck is full notifies the Warehouse Manager to dispatch the truck. The Automated Warehouse Management System that is being tested today is designed for large, high traffic commercial warehouse operations

SRT-2 Automated Warehouse Management System

We are working with a company on issues of automation, as well as being funded by the National Institute of Health for this work. Let me tell you a little about the system you will be helping us test. The company first became involved in sensory technologies in 1975 with the sole mission of creating advanced scanning and decision making systems for warehouse loading and shipping applications. In 1985, the company released an Automated Warehouse Management System called the SRT and in 1997, released a Smart Automated Warehouse Management System, the SRT-1. The company's latest groundbreaking system, the SRT-2, utilizes advanced decision algorithms and sensing technologies that have the ability to adjust to differing warehouse and loading conditions. Testing of the SRT-2 indicates that it is the industry standard for accuracy, reliability, and robustness and is still considered the leader in Automated Warehouse Management systems. Because this is a well proven Automated Warehouse Management System, it is expected that the SRT-2 will perform at a high level with no performance errors.

Standard Expectancy Condition

An Automated Warehouse Management System is a system that scans the inside of truck trailers, calculates the amount of space available in the truck, loads shipments onto the truck, determines if the truck is full, and when the truck is full notifies the Warehouse Manager to dispatch the truck. The Automated Warehouse Management System that is being tested today is designed for large, high traffic commercial warehouse operations

SRT Automated Warehouse Management System

We are working with a company on issues of automation, as well as being funded by the National Institute of Health for this work. Let me tell you a little about the system you will be helping us test. The Automated Warehouse Management System that you will be interacting with today is very reliable but may make errors. Two types of errors can potentially be committed: a false alarm or a miss. A false alarm is when the system indicates that a truck is full when in fact it is not full. For example, like when smoke alarm sounds when there is no fire. A miss is when the system fails to indicate that the truck is full when in fact it IS full. For example, when there is a fire but the smoke alarm does not sound.

5. Please indicate how you perceive the relationship between automated systems and human users.

1	2	3	4	5
Automation works for the human		Collaborative Team		Human works for the automation

Appendix D: Warehouse Description

Warehouse Description

Large Warehouse operations are high paced environments. The presence of multiple loading docks results in numerous trucks being loaded and unloaded at the same time. Full trucks are therefore often being dispatched by the minute. In addition, shipments are received and placed into inventory at an even faster pace. Warehouse managers must oversee and coordinate both the dispatching of full trucks and the receiving of shipments. The cost of dispatching trucks that are not full is enormous; likewise, the cost of overfilling trucks is enormous. In large warehouses, a shipment that is placed into inventory incorrectly can be lost forever, resulting in the cost of lost inventory. To increase revenues and reduce lost shipments, warehouse owners are intensely interested in ways to improve efficiency and productivity of warehouse managers. Automated Warehouse Management Systems provide this potential.

Today you will be playing the role of a warehouse manager involved in usability testing an Automated Warehouse Management System. An Automated Warehouse Management System is a system that scans the inside of truck boxes, calculates the amount of space available in the truck, loads shipments onto the truck, determines if the truck is full, and when the truck is full notifies the Supervising Warehouse Manager to dispatch the truck. The automated warehouse management system that is being tested today is designed for large, high traffic commercial warehouse operations.