DISSERTATION

THE INFLUENCE OF TRUST, SELF-CONFIDENCE AND TASK DIFFICULTY ON AUTOMATION USE

Submitted by

Colleen E. Patton

Department of Psychology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2023

Doctoral Committee:

Advisor: Benjamin Clegg

Christopher Wickens
Gwen Fisher
Francisco Ortega

ABSTRACT

THE INFLUENCE OF TRUST, SELF-CONFIDENCE AND TASK DIFFICULTY ON AUTOMATION USE

Automation can be introduced statically or dynamically to help humans perform tasks. Static automation includes always-present automation types, whereas in dynamic automation, the presence of automation is controlled by another source, typically a human. In static automation, trust, automation accuracy, task difficulty and prior experience with the automation all contribute to the human dependence on the automation. In the dynamic literature however, a small body of research suggests that accuracy and task difficulty do not impact the decision to use automation, but a combination of trust and self-confidence does. The difference between the influence (or lack thereof) of task difficulty in static and dynamic automation is unusual, and prior literature does not make a strong case as to why this difference exists. Through three experiments, the influences of task difficulty, prior experience, trust, self-confidence, and their interactions are investigated.

Experiment 1 used a dual task warehouse management paradigm with a lower-workload and higher-workload version of the task. Results indicated that trust-self-confidence difference was related to automation use, such that higher trust and lower self-confidence led to more use. Additionally, the difficulty manipulation did not have an impact on automation use, but self-confidence did not change across the two levels of difficulty. Experiment 2 investigated four levels of difficulty through a dynamic decision making task with participants detecting hostile ships. There was a difference in automation use at the easiest and most difficult levels, indicating that if the task difficulty difference is salient enough, it may influence automation use. The trust-self-confidence relationship was also present here, but these measures were only collected at the end of the task so their influence across the difficulty levels could not be measured. Experiment 3 used the same paradigm as Experiment 2 to investigate how perceived difficulty, as compared to objective difficulty, influences automation use. Results indicated that perceived workload influenced automation use, as did the change the trust-self-confidence difference.

The findings of these experiments provide insight into how trust and self-confidence interact to influence the choice to use automation and provide novel evidence for the importance of workload in discretionary automation use decisions. This suggests the importance of consideration of human operator perceptions and beliefs about a system and of themselves when considering how often automation will be used. These findings create a foundation for a model of influences on automation use.

ACKNOWLEDGEMENTS

To my advisors, Ben and Chris – thank you. In the past four years, I have grown tremendously as a person and scientist thanks to your mentorship and guidance. Although it has not been an easy road – a pandemic, my own stubbornness, a tendency to email both too often and not enough – I consider myself incredibly fortunate to have been a part of this lab. This document is meant to be a culmination of my scientific knowledge and interests, but I believe if you look closely, you'll see clearly quite a few pieces of the knowledge and interests you have imparted on me. I take these into the "real world" as seeds of new beginnings, encouraging growth both of my own and of my future graduate students. Without either of you, I would not be the scientist I am today.

To Blake, the best computer scientist teammate I could have ever asked for. Thank you for all of your coding, pulling data, and willingness to help no matter how silly my request.

To my family and friends, who fielded panicked calls at all hours of the day and night, who sent me words and tokens of encouragement and congratulations, and who helped entertain my dogs on the long work days, thank you. You kept me going, and gave me the push to finish.

To Brandon, there are not enough words to describe my gratitude. You believed in me when I did not, you were a shoulder to cry on more times than I can count, you celebrated every success bigger than I did, you kept my stomach fed with continuously improving barbeque, you listened to me babble about research ideas and findings, you sat through dozens of practice presentations, and above all else, you continued to love me through every high and low. This dissertation is mine, yes, but it is also yours. You were the sounding board, the support system, and the "trophy husband" who made it possible. I love you. Thank you.

TABLE OF CONTENTS

CHAPTER 1 - TRUST AND DEPENDENCE

Introduction

Much of the human-automation literature is focused on understanding when and how human operators decide to use or depend upon automation. This should be no surprise, as ultimately, understanding and predicting how automation changes or impacts tasks with human involvement is crucial for automation designers in real-world scenarios, especially in safety-critical situations such as the implementation of self-driving vehicle technology. Often, human operators drive accuracy of the human-automation team below the accuracy of the automation itself due to incorrectly identifying accurate automation recommendations (Bartlett & McCarley, 2017, 2021; Boskemper et al, 2022). This finding indicates that humans are not depending on the automation in an ideal or optimal way – an issue that human automation interaction researchers often aim to mitigate.

Simply put, dependence is acting upon a system's recommendation or using automation. It includes both reliance (agreeing with the automation when it suggests there is no target) and compliance (agreeing with the automation when it suggests there is a target) (Meyer, 2004), but can also be used to define automation use in situations where the automation is not a decision aid – for example, self-driving vehicle automation. The definition of dependence contains a key piece for measurement: it is a *behavior*. Therefore, dependence must be measured based on actions or decisions of a human operator. There is an abundance of research on what influences dependence, and a number of variables have been identified as potential influences. One variable that has been extensively studied is trust.

Trust as a construct in human-automation interaction evolved from human-human trust. Although the literature in the area is abundant, there is still not a single agreed upon definition (Sheridan, 2019). One generally accepted defintion is from Lee and See (2004), who define trust as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability." This definition contains a key facet of the concept of trust - the idea of trust as an *attitude*.

In an early paper discussing trust in automation, Sheridan (1989) suggested that dependence is an attribute of user trust. Specifically, he claimed that depending on automation leads to higher trust because we want to trust the things or people that we depend on. This set an early precedent for the pervasive idea that trust and dependence are closely related, and many more recent discussions of trust and dependence often conflate the two, suggesting that dependence can be used to measure trust, as if they have a correlation of 1. Specifically, the importance of 'calibrating trust' is often brought up. This is the idea that when humans align their trust in the system with the system's reliability, they will use the system at the optimal rate. Intuitively, it makes sense that increasing trust will lead to higher dependence, or vice versa, and this is sometimes true. However, when we consider trust as an attitude and dependence as a behavior, it should be quickly clear that the two constructs are not the same. This is often acknowledged by researchers (Merritt et al., 2015; Lee & See, 2004), yet the literature explicitly comparing the two and the independent factors that influence each is limited, and a review of the main findings and implications has not yet been published.

Therefore, this chapter reviews the literature to provided quantitative evidence that trust and dependence are not strongly correlated. While they are not completely unrelated, they also are not closely correlated enough to be used as proxies for each other – an idea that, while discussed among scientists frequently, is often ignored (see Lyons & Stokes, 2012; Monroe & Vangsness, 2022; Rice, 2009; Rice & Geels, 2010; Spain & Bliss, 2009, for examples). This chapter will define dependence and trust in more detail, and then review the current available data to show the separation of trust and dependence.

Operationalizing Dependence

Dependence is a behavioral measure that is used to quantify how often a person is using automation when it is present. For alerting systems or diagnostic aids, dependence can be specified into reliance and compliance (Meyer, 2004). Reliance is when a human agrees with a non-alert or a state of 'all is well' from the automation, regardless of the true state of the task. Reliance is influenced by miss-prone automation, such that a lack of alerts to true states that require alerts tend to influence operators to show less reliance on the automation (Dixon et al., 2007). Reliance is also directly related to complacency, which is when a human operator over relies on the automation (Bahner et al., 2008; Parasuraman & Manzey, 2010). Separately, compliance is when an operator agrees with an alert from automation regardless of the true state. We see compliance degraded with false alarm prone automation, such that automation that sets off a large number of untrue alerts will lead operators to take longer to respond, if they respond at all. Dependence also exists for non-alerting systems, and in these cases would be operationalized through use of the system. For example, dependence on the automation for a self-driving car can be measured through how often the system is activated and when the human is not correcting the automated aspects such as steering or braking.

Dependence is also sometimes measured by eye tracking metrics and visual scanning. Assuming that visual fixations are indicators of attention, metrics such as how often and how long a person looks at the automation can be used to determine their level of dependence on it.

For example, Dixon et al., (2007) found that automation with many misses (lack of an alert to a true-present target) led to more glances over at the automated task when the automation did not provide an alert, indicating that the operators were less dependent on the automation. This idea is returned to later.

Dependence is of particular interest to practitioners and researchers alike, as the level at which a person depends upon or uses the automation when it is present is often predictive of performance outcomes. Over-dependence or under-dependence, also termed ‘misuse’ (Parasuraman & Riley, 1997), can lead to poor performance outcomes due to agreement with the automation when it is incorrect or disagreement when it is correct. Dependence is sometimes calibrated to the reliability of a system (known as ‘probability matching’ [Dorfman, 1969; Bliss et al., 1995]) but often there are still mismatches in correct automation recommendations and correct human responses. This leads to human-automation team performance at or below the automation’s own accuracy levels (Bartlett & McCarley, 2017, 2021; Boskemper et al., 2022). Much research has indicated that one of the strongest influences on dependence is automation reliability or accuracy (Chancey et al., 2017; Dixon et al., 2007; Dixon & Wickens, 2006; Hutchinson et al., 2022), such that as automation becomes more reliable, operators use it more. It is also often thought that trust in the automation strongly contributes to dependence.

Trust in Automation

A full review of trust in automation is outside the scope of this chapter, and readers are encouraged to visit Hoff & Bashir (2015), Lee and See (2004), Chiou and Lee (2023), or Kaplan et al. (2023) for full reviews. Instead here, a brief summary of the findings of these reviews is provided.

Trust is a dynamic process (Yang et al., 2021), it changes over time and across situations. This can be seen in the three main types of trust suggested by Hoff and Bashir (2015): dispositional, situational, and learned. Dispositional trust is an enduring tendency or state of an individual to trust automation regardless of the context. This is typically influenced by factors beyond automation itself, such as culture and age. Situational trust is dependent on the specific context of the interaction with the automation. This can be influenced by external factors – the system, the task it is being used for, risks and benefits – or by internal factors of the operator such as self-confidence, mood, etc. Lastly, learned trust stems from past experiences with a specific system. This can often change from initial trust (dispositional or situational) and is influenced by a large number of factors such as transparency, performance and predictability of the system, and types of automation failures. Similarly, Schaefer et al., (2016) suggested that trust has three properties: trait, state, and dynamic change. These can be closely tied to dispositional, situational, and learned trust, respectively, from Hoff and Bashir (2015).

In practice, trust is measured dozens of different ways - Kohn and colleagues (2021) found over 30 measures of trust in their literature review. Measures span self-report surveys, physiological and behavioral measures. The most cited way to measure trust is through the Jian et al. (2000) scale, which empirically derived a number of questions in a self-report survey designed to capture trust in a specific system. Physiological measures, including neurological measures, galvanic skin response, heart rate variability, and eye tracking, are also becoming popular in trust measurement. Using eye tracking, the frequency with which an operator looks at the automation or looks at the raw data that the automation is using is assumed to be an indicator of trust, such that more monitoring indicates less trust. The confound of trust and dependence comes from behavioral measures of trust, where researchers use compliance and reliance to indicate trust levels.

Simply from the definitions and operationalizations of trust and dependence, it should be apparent quickly that saying trust and dependence are the same construct is incorrect. Trust is an attitude and hence a subjective state, whereas dependence is an outcome or a behavior. Additionally, all three types of trust can differ from each other within one person and one context. While trust and dependence are not mutually exclusive, to use the terms interchangeably, or use dependence as a measure of trust (or vice versa) can often lead to incorrect conclusions. Lee and See (2004) suggest "that trust guides, but does not determine, reliance", and more recent literature supports this idea.

Workload

If trust guides, but does not determine, reliance, there must be other factors influencing the trust-dependence relationship. One omnipresent factor in human-automation interaction is workload. Workload lacks a unanimous definition (Longo et al., 2022), but there is some consistency across definitions in that workload always contains a task and a subject. One definition that has gained some traction is the idea that workload is a ratio of the mental resources required to the mental resources available (Carswell et al., 2005). This is not the same as task difficulty, although sometimes the two are related, especially in what is termed "underload" situations, where the mental resources of an operator are not at capacity. When resources are at capacity, however, this is termed the "red line" of workload (Grier et al., 2008).

Workload, or the amount of resources used, can be measured objectively or subjectively. Importantly, there are four types of workload to differentiate between. Objectively, workload or task load can be measured by the complexity of the task or its bandwidth. For example, increasing the number of items an operator must track will inherently increase workload. It is worth noting that this may not be a precise measure of workload, but will generally indicate differences that correlate with subjective workload between tasks. Another way to measure workload objectively is to use primary task performance. This is a valid measure when resource use is over the "red line" in the overload condition. However, if additional resources are still available when under the red line, performance will be at ceiling and cannot reliability indicate the amount of resources being used. If resources are at capacity though, as workload increases, performance will decrease. This is in line with the idea of resource and data-limited tasks (Norman & Bobrow, 1979). Subjectively, workload can be measured through questionnaires like the NASA-TLX (Hart & Staveland, 1988) or the ISA (Tattersall & Foord, 1996). In this way, participants are asked to report aspects of the task demands, such as how they perceive their performance or how much effort they have to exert. In theory, these subjective measures should align with objective measures, although they may not always when resources are at capacity. These subjective measures can be useful to understand how an individual perceives the workload of a task, especially below the red line of workload.

Aside from explicit measurement of workload, it is important to differentiate between primary task load and concurrent task load. This is apparent in most real-world situations and multi-tasking experimental design. Workload or difficulty changes can be imposed on the primary task, or task of interest, but can also be affected by other environmental changes or secondary task changes. For example, if a secondary task become more difficult, the overall workload will increase even if the primary task remains unchanged, as long as the secondary task is being attended to in some capacity (i.e. not completely ignored). Similarly, environmental demands, rather than a secondary task, can impose changes to workload without directly changing the primary task. For example, the task of driving can remain the same, but the concurrent task load can differ if the driving is on a highway during a sunny day with no traffic as compared to driving on a foggy evening with a lot of traffic. This distinction is crucial in the current dissertation's Experiments 1 versus Experiments 2 and 3.

Across this dissertation, these types of workload will be differentiated by using the terms task load (complexity of the task or bandwidth), objective workload (primary task performance), subjective workload (survey-based measures), and concurrent workload (secondary task or environmental factors). In the existing literature, the type of workload is not always explicitly defined, but can be inferred. Typically, issues in automation dependency arise when attentional resources are at or near the point of overload, especially for the task that the automation is assisting. As both primary and secondary tasks become more difficult and require more cognitive resources, dependence on automation tends to increase (Biros et al., 2004; Dixon et al., 2007; McBride et al., 2011; Xu et al., 2007). There is some evidence to suggest the opposite is true (Tikhomirov et al., 2022; Monroe & Vangsness, 2022) but more often as workload or task difficulty increases, humans tend to exhibit automation bias, where they believe the automation is good enough or better than they are, and they no longer need to monitor it closely (Parasuraman & Manzey, 2010; Skitka et al., 2000). This would suggest that when a high workload situation involves an automated aid, it is more likely to be depended on, regardless of trust because when a task is very difficult, the operator will almost be forced to depend on the automation in order to continue performing at an acceptable level. This hypothesis is examined in this review.

Automation Reliability

Another metric of interest is automation reliability, or the accuracy of an automated system. The literature seems to suggest that reliability influences dependence. Sometimes, dependence and reliability are almost perfectly matched, called "probability matching" (Dorfman, 1969). This is seen when an operator's reliance or compliance rate is the same as the automation's reliability. More often though, dependence is not perfectly matched to the automation's reliability because users underestimate its actual reliability (e.g., Hutchinson et al., 2022; Patton et al., *in press*; Pop et al., 2015; Wiegmann et al., 2001). Although seldom calibrated, people are sensitive to automation accuracy (Hutchinson et al., 2022; Madhavan & Wiegmann, 2007; Patton et al., *in press*). This means that, for example, operators will use an 80% reliable system more often than a 60% reliable system, but use will not match reliability exactly.

While it seems clear that reliability and dependence are linked in some way, there are also suggestions in the literature that trust and reliance are linked, such that as reliability increases, trust also increases (Dzindolet et al., 2003; Hoff & Bashir, 2015; Ross et al., 2008). However, the connection between trust and dependence, as influenced by reliability, has not been quantified.

Current Review

Although there is a general agreement that trust and dependence are not the same, this concept has not yet been quantified and the literature continues to conflate the two (i.e. Monroe & Vangsness, 2022; Rice, 2009). Hoff and Bashir (2015) provide the insightful theory that trust and reliance are less related when operators are unable to judge how much the automation is helping them, and when there is lower decisional freedom, and suggest that additional situational factors, such as the level of effort required to engage a system, the alternatives to using automation, time constraints, and the operator's situational awareness and physical well-being, can impact dependence without changing trust. Yet, quantified metrics of the relationship between the two have only been examined individually, and in a handful of studies. The measured difference between these two constructs could range a large scale, and so without a review of the literature and effects, the implications of using dependence as a metric for trust is unclear.

Of specific interest here is also the way that workload and reliability play into the trust-dependence relationship. Although previous trust reviews suggest a number of factors that influence dependence and trust, workload and automation reliability are of interest because they are involved in every automation interaction. Both variables are always able to be measured or quantified (even if they are not explicitly measured), and often are of interest in real world environments. Therefore, their influence on the trust-dependence relationship is worth consideration across most contexts.

The current chapter aims to provide a quantification through correlation of the way in which trust and dependence are related. From published papers regarding trust, dependence, compliance, and reliance in the Human Factors field, the reported correlations of participant trust and dependence were collected and averaged to understand the relationship. It was expected that trust and dependence overall would have a moderate correlation due to a wide range of correlations observed between the two variables. Further, the way in which workload and reliance interact with the trust-dependence relationship was of interest.

Methods

Published, peer-reviewed articles were collected from Google Scholar, the Human Factors Journal, and the Human Factors and Ergonomics Conference Annual Proceedings. Search terms for all Human Factors journals were “trust + dependence”, “trust + compliance” and “trust + reliance”. These search terms were then repeated with the addition of “ + workload” and then “+ reliability”. In Google Scholar, these search terms had “automation” added so that human-human trust would not be included. This allowed an encompassing search of papers that covered trust and dependence generally, as well as with workload and automation reliability manipulations.

Articles were then reviewed to ensure they measured both subjective trust and objective dependence. Subjective trust could be measured through any type of questionnaire and at any point in a study. Although the differentiation between types of trust is covered in review papers, few experimental papers specify the type of trust they measured and therefore, those that did were included as separate measures of trust but were not correlated any differently. Dependence could be measured as automation use (turning the automation on) or agreement with the automation. Importantly, dependence could not be gathered from accuracy reports and thus papers that did not report reliance, compliance, or use as their form of dependence were excluded. Additionally, one form of dependence that appeared in multiple articles was monitoring, or looking at the automation or additional data from the automation. This was considered a separate form of dependence from reliance and compliance. Articles containing a reliability or workload manipulation were also categorized. This created three categories of articles (see Table 1 and Appendices B and C for reference lists) - trust and dependence generally (14 articles), trust and dependence with a workload manipulation (4 articles), and trust and dependence with a reliability manipulation (6 articles).

Where possible, correlations between trust and dependence were gathered. Studies that manipulated workload or reliability typically reported trust and dependence separately. Most of these studies also did not report means of both metrics, so specific correlations could not be gathered. Instead, general trends were investigated (i.e., increases in both metrics across conditions. If a paper did not report both metrics, it was not included in this review.

Results

Trust and Dependence

Correlations that were reported between participants for subjective trust and objective dependence – operationalized as automation use, reliance, or compliance – were averaged. The raw numbers and associated studies are provided in Table 1. An average weighted correlation between trust and dependence was computed, resulting in a mean r value of .15. The range of correlations was from -.23 up to .71, with a standard deviation of .27. This small correlation would suggest that, on average, trust accounts for only about 2% of the variance in dependence across these studies. It is worth noting that the Wiczorek and Meyer (2019) paper reports a significant correlation between trust and reliance, but then does not report information for a non-significant correlation with compliance, which may skew the average correlation slightly.

The r values were turned in Cohen's d effect sizes using the formula $r = \frac{d}{\sqrt{(d^2+4)}}$. The average effect size was d = 0.33, with a standard deviation of 0.63. This aligns with the correlation finding, suggesting that trust is somewhat related to dependence, although the effect is not large.

Table 1. Correlations for trust and various operationalizations of dependence. DOA stands for degree of automation.

Correlation	Number of Subjects	Correlated With	Citation

0.3	360	lower DOA, compliance	Chien et al., 2020
0.09	360	higher DOA, compliance	Chien et al., 2020
0.27	120	compliance	Chien et al., 2020
0.17	120	compliance	Chien et al., 2020
0.44	120	compliance	Chien et al., 2020
-0.06	100	compliance	Davenport & Bustamante, 2010
0.17	100	reliance	Davenport & Bustamante, 2010
0.15	25	reliance	Wang et al., 2011
0.47	26	reliance	Wickens et al., 2020
-0.11	58	compliance	Huegli et al., 2022
0.14	58	reliance	Huegli et al., 2022
-0.22	98	compliance	Huegli et al., 2022
-0.03	98	compliance	Huegli et al., 2022
0.2	98	reliance	Huegli et al., 2022
0.13	98	reliance	Huegli et al., 2022
-0.16	6	use	Muir & Moray, 1996
0.71	6	use	Muir & Moray, 1996
-0.23	18	use	Lee & Moray, 1992
-0.18	18	use	Lee & Moray, 1992
-0.17	18	use	Lee & Moray, 1992
0.47	23	use	Tenendfeld et al., 2020
0.11	26	reliance	Wang et al., 2011
0.21	26	reliance	Wang et al., 2011
0.31	80	reliance	Wiczorek & Meyer, 2019
0.45	72	compliance	Pharmer et al., 2021
-0.25	35	dispositional + monitoring	Hergeth et al., 2016
-0.27	35	situational + monitoring	Hergeth et al., 2016
0.35	45	peeking	Pak et al., 2012
-0.41	6	monitoring	Muir & Moray, 1996

Trust and Monitoring

The literature that reported correlations between subjective trust and monitoring of the automation, as inferred from eye movements, appeared fundamentally different from the trust and dependence literature, so these correlations were averaged separately. Across four studies, the average correlation was r = -.14. However, monitoring in the Pak et al., (2012) study was not based on eye movements, but instead allowed participants to seek additional information before deciding to accept or reject the automation's suggestion. This correlation could be considered a different type of relationship, due to the change in effort. Choosing to click on something to reveal more information that must then be integrated and understood is quite effortful, whereas glancing over at a system and doing a quick, heuristic check of the status is much less effortful. Therefore, the average of monitoring-trust correlations was repeated without this study. In this case, r = -.31, which is a stronger correlation and suggests that trust explains about 9.6% of the variance in monitoring (eye movements or dwell times): those who trust more, monitor less.

Workload

Whereas the previous correlations were entirely between subjects, there is also an interest in the way the trust-dependence relationship changes across conditions. Therefore, studies that varied workload were investigated. One study was removed from review because the workload manipulation produced no difference in performance or subjective workload reports, indicating a null effect of workload and thus rendering the findings unrelated to workload changes. Two studies varied workload on the concurrent task and one study varied workload of the automated task. Correlations between trust and dependence across different workload levels and summary statistics were not reported often enough for numerical relationships between trust and dependence to be analyzed. However, general trends across studies were collected, specifically in the way trust and dependence changed between high and low workload conditions.

Overall, the trends were inconclusive, which comes as a surprise. Chien et al. (2019) found higher trust and compliance in the higher task load condition than in the lower task load condition, but reliance was highest in the low task load condition. Zhang & Yang (2017) reported no effect of task load on reliance or trust between conditions, although dwell time on the automated task decreased as task load increased. This suggests a dissociation, as trust did not change from low to high workload, but monitoring decreased. This is in line with the negative correlation found between trust and monitoring overall, as reported above. Karpinsky et al. (2018) also found less dwell time on the automated task (i.e. less monitoring) as subjective workload increased, but trust decreased. Together these three studies provide evidence for both the association and dissociation of trust and dependence in high workload contexts, but without quantifiable evidence, the exact relationship cannot be concluded. The implications of these findings are expanded upon in the Discussion.

Automation Reliability

Between-condition correlations of trust and dependence across varying levels of automation reliability were also not reported consistently enough for numerical relationships to be analyzed. Instead, reports of changes in dependence and trust were recorded and analyzed for general trends. For example, in Du et al., (2020), trust, compliance and reliance were all reported individually to have increased with increasing automation reliability, but the three metrics were not reported in relation to each other. However, the literature here had consistent trends for both trust and dependence to increase as reliability increased, and vice versa (trust and dependence decreased as reliability decreased). This was shown in one study to be true even when reliability perceptions, rather than objective reliability, was manipulated (Bowden et al., 2021).

While these trends do not provide evidence for the amount of variability in dependence explained by trust or automation reliability, a number of studies have investigated the mediation model of these variables. It seems that trust mediates the relationship between reliability and dependence (Chancey et al., 2015; 2017; Hussein et al., 2020). This suggests that trust can help explain the relationship between dependence and reliability, but does not allow strong conclusions about the relationship of trust and dependence to be made.

Discussion

This review set out to understand the extent to which trust in and dependence on automation are related. While many researchers will agree that trust and dependence are not the same, the list of published studies that use dependence as a measure of trust is quite extensive, indicating that the idea of trust and dependence being separate constructs is not always acknowledged in experimental design and data collection. However, the results of this review indicated that, as expected, trust and dependence are not strongly correlated with an overall r-value of .13. This suggests that using dependence to measure trust is unlikely to provide an accurate measure of subjective trust, and vice-versa. Separately, trust and monitoring of the automation, operationalized through dwell time or scanning patterns, were moderately (negatively) correlated.

Dependence and monitoring are not the same construct, but this relationship – less monitoring with more trust - is noteworthy for two reasons. First, there is a need for a dynamic measure of trust. Subjective trust is often measured “dynamically” through questionnaires that are presented at the end of an experiment (i.e. Yang et al., 2021), or throughout an experiment by pausing the trial and presenting the questions (i.e., Foroughi et al., 2021). These may not capture the way trust can change moment to moment, and because dependence provides a view over time of how operators or participants are using the automation, it intuitively makes sense that dependence might be a better way to measure trust overall. However, the current results suggest that is often not the case. Instead, monitoring of the automation may be the best behavioral measure of trust. With an average correlation of r = -.31, there is a moderate relationship to suggest that more visual attention to the automation indicates less trust. This would provide a non-intrusive way to measure trust over time, although with only about 10% of the variance explained, there is still room for changes in trust that are not caught through monitoring.

The second reason this relationship is noteworthy is from a system designer's point of view. Designers typically are primarily interested in use of a system. Knowing that there is some relationship between monitoring and trust, a designer could predict a user's trust from their interaction with the system, allowing them to gather both use and trust information from a single metric.

While there are cases where the correlation is quite large (i.e., Muir & Moray, 1996; Experiment 2, r=.71), the overall relationship between trust and dependence is small, which suggests that other variable(s) are likely changing trust or dependence, but not both. Workload and automation reliability have both been thought to drive changes in dependence and trust individually, but the studies that explicitly investigated these variables showed that workload may not have as strong an impact on the relation between trust and dependence as anticipated. Instead, the findings across these studies were mixed. This is not to say that workload has no impact on trust or dependence, as all of the studies included in this review have found effects of manipulated workload on both variables individually. However, when these variables are considered independently and summary statistics are not provided, the ability to understand how workload interacts with both variables is limited. This highlights the need for researchers to report trust and dependence numbers *and* relationships in their work, especially surrounding workload and automation use.

In contrast to workload, automation reliability, however, did show a consistent trend such that both trust and dependence increased as reliability increased. Although again this trend in the relationship between the two variables was not able to be quantified, the consistency across the few studies may be, in part, responsible for the pervasive idea that trust can be measured through dependence. However, it is important to note that in these cases, while trust and dependence increase in tandem, they are not the same metric. Rather, these studies suggest that if an experiment manipulates automation reliability and reveals a change in dependence (or trust), it may be assumed that trust (or dependence) is changing in that same direction. It does not allow researchers to quantify the *level* of trust or dependence without explicitly measuring either variable.

Although automation reliability and workload provide some insight into influencing factors, there are of course other factors not addressed here that must influence the trust-dependence relationship. Hoff and Bashir (2015) cover many of these in detail, differentiating between influences of the human operator, the environment, and the automated system. Reliability and workload are system and environmental variables, respectively, but the human variables and many other factors from the system and environment have not been researched in enough detail to provide an understanding of when trust and dependence are strongly correlated.

Future Directions

This literature review makes clear a need for four main considerations in future research:

1. There is a need for reporting subjective trust *and* objective dependence measures more often, especially if claims are made about both metrics. The literature on trust and dependence individually is vast, but the number of studies that provide results of both is very limited. While studies that only investigate trust and dependence

correlations may not be likely candidates for publication in prestigious journals due to their “intuitive” nature, their importance should not be overlooked. These findings should also be reported in papers who may not be solely focused on the relationship but have measured both outcomes for other purposes.

2. Relatedly, there is a need to report the results of studies in ways that allow for meta-analyses and reviews to be conducted. While this is not limited to the trust and dependence literature, it is one area that makes clear the benefit of reporting means, standard deviations, and effect sizes for all effects in an empirical research paper. Although the literature that explicitly reports both trust and dependence is already small, many papers had to be omitted from this review due to a lack of reporting of summary statistics.
3. The idea that “trust calibration” is a key piece of human-automation interaction should be considered carefully. Trust calibration is a term that refers to the idea that an operator’s trust aligns with the reliability of the automation (Lee & See, 2004), and implicitly suggests that their dependence or use of the automation will be at the proper level. However, when trust and dependence are not strongly correlated, the calibration of trust becomes less important if it is being used as a metric for dependence. Rather, trust should be calibrated to reduce the negative impact on the operator’s experience or perception of the automation and task. This is because trust, regardless of dependence, impacts how users perceive the system and task. Consider, for example, an operator in a high workload environment. If they do not trust the system but must depend on it, this can create dissonance, which can cause additional stress and may increase monitoring of the automation. This can hurt performance of a

non-automated task performed concurrently due to a decrease in attention on that task (similar to the effects of false-alarm prone systems in Dixon et al., (2007)). Alternatively, if an operator trusts a system and needs to rely on it, this can lessen their workload and stress, and allow them to focus on the non-automated task to improve performance. In this way, two operators could be performing the same task with the same automation, but have very different perceptions of the stress and difficulty of the task based on their trust in the automation.

4. Systematic, empirical investigations are needed to better understand the mechanisms through which trust and dependence are related. This relates to the need for reporting these measures directly, but also would allow for scientists and practitioners alike to better predict when trust and dependence will be strongly correlated. Many of the studies that report trust and dependence lack reporting of other outcomes or variables such as task performance, training, risk, or perceptions of the context or automation. Rather, the novelty of each subsequent study seems to stem from slight changes in the context of the experiment, more precise methods, or a different population. There is a need to investigate other manipulations within human, contextual and automation-centered factors while reporting subjective trust and objective dependence measures.

Conclusions

It likely comes as no surprise to most automation researchers that trust and dependence were weakly correlated, and that monitoring and dependence were moderately correlated. Yet, the literature that measures trust through dependence is vast enough to suggest that a quantitative analysis of their relationship was necessary. Trust and dependence may only have a small correlation, but that is not to say they are entirely independent of each other. Rather, care must be taken by researchers to determine the context in which the two constructs are being measured and the relative importance of each in the applications of the research. Both trust and dependence have implications for design and use of automation through their impacts on performance and perception of the human operator. Yet, by measuring one as the other, the important differences between the constructs are lost.

While more research is needed to better validate the relationship between trust and dependence, it seems from the current literature that the influences on dependence are much more complex than only trust. This brings forth a new question: if trust is not the main factor in automation dependence, what is? Furthermore, dependence is not the only metric of automation interaction. There is also an active choice to use automation – one that might differ from the factors influencing dependence entirely.

CHAPTER 2 – INTRODUCTION

Introduction

Automation is becoming ubiquitous. If we consider automation a tool that completes tasks that humans used to complete (Parasuraman & Riley, 1997), we can see its presence in our alarm clocks, email scanning algorithms, phones, and more. As automation becomes more commonplace, the questions of how people interact with it and how it changes a situation become more important. Automation is seldom presented in a non-negotiable way – human operators can typically choose to listen to or ignore the automation, or turn it on or off. When this happens, it may lead to a benefit in performance – when an important email document gets sent to the spam folder and the user reads and responds to that email, that benefits the task. Other times, ignoring the automation can result in serious consequences – not leaving a building when the fire alarm indicates a fire would be disastrous if the fire really is present.

While research on human-automation interaction has existed for decades, much of it focuses on automation that is always present, even though there are numerous situations in which the presence of automation can be controlled by a human. There is a need to understand how the conclusions from the always-present automation translate to the human-controlled automation, but also a need to understand what influences the human decisions to use automation at all. Chapter 1 reviewed the way that trust and dependence – a form of automation use – are related, and briefly discussed workload and automation reliability. The small average correlation between the two constructs indicates that there are other factors playing a role in the choice to depend upon, or use, automation.

This dissertation will briefly review the factors that influence dependence on always-present automation. Then, the existing literature on human-controlled automation (adaptable and discretionary) will be reviewed. The current experiments aim to better understand how trust and dependence are related (specifically, the impact of considering trust and self-confidence in conjunction on automation use) as well as the impact of prior experience with automation and task difficulty on automation use.

Prior Experience

Intuitively, prior experience with automation would be thought to have an impact on use. A simple example of this is older generations, who often prefer non-automated items such as a physical newspaper, over younger generations who prefer automated items such as a news app showing the most relevant stories. This is often attributed to growing up with/without technologies and therefore having difference levels of experience with the item, creating almost an anchoring effect for that item.

Research has shown that familiarity with an automated driving system can increase trust and subsequently better use of the automation (Hergeth et al., 2016; Koustanaï et al., 2012). In a non-driving domain, Gefen et al., (2003) found that people who had previously used an e-vendor (i.e. online store like Amazon) trusted it more and perceived the vendor to be more useful, whereas new customers only guided their use by their trust and not perceived usefulness. It is worth noting that while there is literature on domain experts and automation use (i.e., Chavaillaz et al., 2019; Nourani et al., 2020), the focus here is on familiarity with a specific system, rather than familiarity in the form of domain knowledge.

The literature here is limited, but it seems that prior experience with automation, as well as the sequence in which automation and manual control are experienced, may play a role in automation use. This is seen in some literature in the form of order effects. Moray et al., (2000) found that operators felt more self-confident in their own performance if they performed the task manually before doing it with the automation than vice versa. Similarly, de Vries et al., (2003) found that the trust-self-confidence difference (trust minus self-confidence) was lower when participants performed a task with automation first rather than completing it manually first. These findings suggest that automation use would differ between the groups, based on previous literature suggesting that automation use is greater to the extent that trust outweighs self-confidence (detailed below). Typically, these order effects are attempted to be controlled for in experiments but here, they are of interest as a way to understand the potential anchoring effects of using automation or experiencing a task manually as the first introduction to it.

Unaided Operator Accuracy

One other variable that intuitively would seem to influence automation use is unaided operator accuracy in a task – if an operator is performing poorly on their own, they should want the help of automation. This links back to the findings that dependency increases with task difficulty, regardless of trust (Biros et al., 2004; Dixon et al., 2007; McBride et al., 2011; Xu et al., 2007). However, little evidence has been published on how human accuracy influences automation use, but one study found no correlation between manual accuracy and automation use (Riley, 1995). This may be for two reasons. First, some reviews of the literature have suggested that operators struggle to know when to use automation if it is not clear how much it is actually helping them (Hoff & Bashir, 2015), or when the state of the automation (i.e., failure) is not clear (Riley, 1995). Additionally, human operators may struggle to accurately identify their competency in a task. Literature has shown that humans are often overconfident in their abilities, especially in difficult tasks (Herdener et al., 2016; Horrey et al., 2009), which could lead to incorrect perceptions of the need for automation. Part of the issue in these circumstances may be that humans are not attending to the implications of the presence of automation, or the automation failures are not salient enough. One way to combat these issues is to have humans more “in the loop”.

Summary

Trust is often thought to be a strong indicator of dependence, and it has been sometimes shown to correlate closely with measures of dependence. It also tends to be a common thread throughout the other influences on dependence, which reiterates its overall importance. However, there are many instances where trust alone cannot perfectly predict automation use, in fact, the average correlation between trust and dependence suggests that trust seems to predict very little about dependence. At this point, four other influences have been briefly reviewed: workload, prior experience, automation reliability, and human accuracy (Figure 1). These influences appear to have a positive influence on dependence – as they increase, dependence increases.

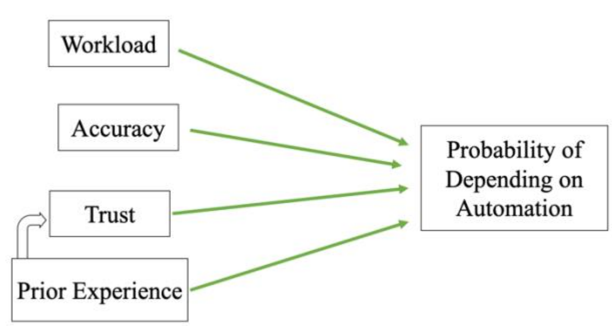

Figure 1. Model of the influences on using automation, including the direction of their effects. The arrow between trust and prior experience indicates the influence prior experience can have on trust.

While the literature reviewed so far appears to provide clear trends of the way that these variables influence automation use, most of this literature has focused only on static automation, or automation that is always present. This creates an issue, as real-world scenarios often allow operators to turn the automation on or off entirely, or choose the level of automation they want at a given point in time. Dynamic forms of automation, where the operator has control over the presence of automation, help keep operators more involved, or “in the loop” in the automated task, which can help prevent complacency issues. However, whether or not the findings from static automation hold in dynamic automation has not yet been heavily investigated. Additionally, some of the findings from the dynamic automation literature may provide insight into why influences like trust are not straightforward predictors of dependence.

Adaptable Automation

The idea of humans “in the loop” or “on the loop” (see Merat et al., 2019 for review of definitions) assumes that humans are involved in the process of which the automation is helping, or are actively monitoring the automation itself. Researchers have found that introducing adaptable automation can improve human in the loop performance (Kaber & Endsley, 2004; Sauer et al., 2012). Adaptable automation is seen when users can choose the level of automation they want at a given time, including turning it on or off entirely. Having a choice of a navigation system providing turn-by-turn directions to a destination or only alerting you of traffic jams is an example of adaptable automation. The research on adaptable automation suggests that it is useful in calibrating trust, increasing acceptance of automation, and promoting faster failure recovery (Calhoun, 2022). However, it also can create a higher workload when the operator is asked to control or choose the level of automation. Likely as a result, people tend to choose one level of automation and stick with it (Chavaillaz et al., 2016) rather than switching across levels.

Adaptable automation also typically shows better performance metrics than adaptive (levels of automation changing based on set metrics such as performance level) or static (always present) automation (Calhoun, 2022). However, research suggests that operators who change levels of automation more often also usually have worse performance (Sauer & Chavaillaz, 2018). While the evidence for adaptable automation is not all negative, it seems that allowing people to choose levels of automation is not a guarantee to improving performance. The issue here could be the choices involved. People are generally cognitive misers and also struggle to understand the cost-benefit analysis of using or not using automation. When providing a number of choices that require effort to implement without training, people may not be willing or able to choose the best option. This idea was supported by a study that trained participants for seven hours and then gave them various levels of automation to choose from (Harris et al., 1995). The trained participants changed levels more often than those who were new to the task, in part because the switch was easier cognitively. However, it is not always feasible to train people for long periods of time. Instead, it may make sense to provide adaptable automation through a single level of automation that operators can turn on or off as they see fit – termed here "*discretionary automation*."

Discretionary Automation

The idea of discretionary automation does not just follow from adaptable automation theoretically, but it exists many real world scenarios. From everyday items such as choosing to turn on an alarm clock or an electronic reminder, to more high-profile cases like self-driving cars, people are often given the option to use or not use automation. There is a small body of research in this area – specifically, the seminal studies of Riley (1995), Lee and Moray (1992), and Muir and Moray (1996) began to study the influences on the choice to turn automation on or off. Beyond these three studies, the literature is sparse. While these studies provide some evidence for the impact of trust, self-confidence and task difficulty on automation use, little research has been conducted in the past two decades. Instead, the focus has been on static and adaptable automation. Yet, as the choice to use automation becomes more commonplace, the importance of understanding the influences on this choice is paramount. The small body of literature that has investigated discretionary automation suggests that some of the key variables in the choice to use automation differ from those that impact dependence on static automation. This sets up a crucial turning point to investigate. If the choice to *use* automation differs in the choice to *depend* upon automation, it creates a massive gap in the literature that reduces the ability of practitioners and designers to best support human operators. A few influences on the choice to use automation have been investigated thus far.

Self-Confidence

As mentioned earlier, the divergence of dependence and trust tend in difficult task scenarios suggests that another factor is influencing the choice to use automation. The unstable relationship between trust and dependence in static automation seems somewhat similar in discretionary automation. Riley (1995) and Lee and Moray (1992) both suggested that trust in automation may drive some of the decision to use it, but it is not the main factor. Instead, they implicate the importance of self-confidence. If a person is highly confident that they can perform a task manually, they are less likely to use the automation. Other studies have continued to show a relationship between self-confidence and use. Riley (1995) found that subjective levels of self-confidence influence automation use, although the subjective levels were not an accurate measure of actual competence at the task. While self-confidence alone seems to be a strong predictor of use, combining it with trust is thought to create an even stronger predictor. If an operator has more confidence in their skills to do the task manually than they have trust in the system, they are less likely to use the automation than when the reverse is true. Riley (1995) and Lee and Moray (1992) both found this to be true, as did De Vries et al., (2003) in a separate paradigm. The current research continues to investigate this measure.

Interestingly, a more recent study found that people who had lower trust in the automation than confidence in their own abilities did not exclusively complete the task manually (Wiczorek & Meyer, 2019), as would have been predicted from the Riley and Lee and Moray studies. However, this is likely the case due to individual differences and does not disprove the self-confidence-trust hypothesis, but rather indicates that it is not a perfect predictor. Self-confidence and trust together may be strong predictors of dependence because trust is affected by system reliability but self-confidence is not (Moray et al., 2000). When combining these two variables, it captures both the system and human impacts on automation use.

Accuracy

Automation accuracy or reliability is thought to be an important component in dependence on static automation (reviewed above), such that as automation becomes better at the task, people tend to rely on the automation more often. In discretionary automation, this finding was not as clear. Chavaillaz et al. (2016) found that the reliability of the automation did not impact the level of automation chosen. While they explicitly investigated adaptable automation, one level was “no automation”, therefore including discretionary automation. In other studies, the reliability of the automated system impacted complacency (defined as overreliance on the automation), but not the choice of which level of automation operators preferred, including manual choice (Chavaillaz & Sauer, 2017; Sauer & Chavaillaz, 2018; Chavaillaz et al., 2019). De Vries et al. (2003), however, found that participants chose to use automation more when it was more reliable. Moray et al. (2000) found that reliability only influenced trust, not self-confidence, which may explain the variation in other findings, should the trust-self-confidence difference be the main influence on use. In all, these findings suggest that the impact of system accuracy or reliability on the decision to rely on automation – which is typically considered automation use in static automation contexts – is different from its impact on the choice to turn automation on or off.

Task Difficulty

Another place where there seems to be a disconnect between static and discretionary automation is in task difficulty. Riley (1994) and Lee and Moray (1992) found that in a discretionary automation task, the difficulty of the task does not seem to impact the decision to use automation. This is also true if task difficulty is measured by human accuracy. De Vries and colleagues (2003) had participants complete a route planning task with an aid that they could choose to use. Participants who performed worse in the manual condition were not more likely to use the automation than those who performed well manually. This is in contrast to the literature on dependence on static automation, as more difficult tasks tend to lead to more dependence (Biros et al., 2004; Dixon et al., 2007; McBride et al., 2011; Xu et al., 2007). It is possible that this is related to self-confidence, where people's actual performance dissociates from their perceived abilities or performance, thereby influencing their automation use due to perceptions rather than objective metrics of performance. This idea, and the disconnect between the impact of task difficulty on static automation dependence and discretionary automation use is of interest in this dissertation.

Summary

Dependence on static automation is influenced by workload, trust in the automation, prior experience with the automation, and accuracy of the automation. These influences have positive effects on dependence and have been shown across a number of paradigms. While static automation is seen in real-world environments, human operators are increasingly often presented with choices surrounding the presence or level of automation. For that reason, research has begun to investigate human-automation interaction with dynamic automation such as adaptable and discretionary automation. The findings in this literature suggest that workload and automation accuracy may not have an influence on the choice to use automation. Instead, self-confidence in completing the task manually seems to exert a large influence, even if self-confidence is not indicative of actual performance.

Current Experiments

The literature on automation use – whether defined as an active choice or dependence – is abundant. However, the influences on dependence versus active choice of use seem to differ in some ways – specifically in the influence of automation accuracy and task workload. The previous studies have created a foundation for future research into discretionary automation, but some of them lacked statistical power, manipulated multiple variables at one time, or have not been replicated in any other paradigms or variations. Importantly, the role of trust in use decisions still needs to be better understood, and the way in which self-confidence interacts with trust needs to be verified. The impact of task difficulty on discretionary automation use as compared to static automation dependence needs to be clarified; specifically, the subjective nature of trust and self-confidence implores the question of other subjective influences on automation use, such as perceived, rather than objective, difficulty. Lastly, I am interested in how prior experience with automation may change the decision to use it. The current research aims to understand if the prior findings surrounding self-confidence and trust, task difficulty, and prior use of automation hold under stronger methodology, and to increase our understanding of the differences between how task difficulty influences automation use in discretionary environments versus its influence in static environments. Four main research questions will be answered across three experiments:

1. To what extent does trust impact the decision to use automation, independent of workload or self-confidence?
2. To what extent does the relationship between trust and self-confidence impact the decision to use automation, and if so, is this a better predictor than trust alone?
3. To what extent does objective task difficulty influence the choice to use automation, or is the perception of task difficulty a stronger influence?
4. To what extent does the order in which an operator completes a task manually versus with automation impact a later decision to use automation?

Two paradigms are used in the current experiments. Both paradigms introduce automation that assists participants in making a decision but provides no action implementation. Participants completed the tasks manually and with automation before being given choices to use automation. Experiment 1 aimed to replicate the finding from Lee and Moray (1992) that higher trust and lower self-confidence would lead to more automation use, and see if task difficulty or workload truly does not impact the decision to use automation. The current study had higher power than previous studies with a bigger sample size and more controlled conditions. The experiment also investigated whether the order in which a person experiences automation – manual task then automated task, or vice versa – would impact the choice to use automation. Experiment 2 investigated objective task difficulty and again the trust/self-confidence relationship in a different context. Experiment 3 aimed to better understand the role of task difficulty in automation use by manipulating perceived task difficulty without changing objective task difficulty.

CHAPTER 3 - EXPERIMENT 1 (Trust, Self-Confidence and Prior Experience Impacts on Automation Use)

Introduction

Previous studies have suggested that task difficulty or workload does not influence the choice to use automation (Riley, 1995; Harris et al., 1995). However, Riley (1995) manipulated many variables at once, making it hard to parse out the effects of each variable individually. Harris and colleagues (1995) also found no impact of task difficulty, but with 10 participants, power was too low to make any strong conclusions. These null findings are somewhat suspect, considering the abundance of literature showing higher dependency when a task is more difficult in static automation contexts (Biros et al., 2004; Dixon et al., 2007; McBride et al., 2011; Xu et al., 2007). The other issue with past research is the conflation of task difficulty and workload. While the two constructs sometimes can be used interchangeably, they are not always equivalent. Part of this difference comes from task difficulty being a manipulated variable, whereas workload is sometimes manipulated and other times measured as a dependent variable (i.e., the perception of workload measured in a post-task survey). It can also differ at certain points of task difficulty, particularly when the task uses all available cognitive resources (see Yeh & Wickens, 1988, for review). Here, task difficulty was manipulated, but a subjective measure of workload was introduced to ensure that the difference could be captured, if necessary.

One of the other unclear pieces of what contributes to automation use is prior experience with the automation. While there is some evidence to suggest that prior experience influences factors like trust, the literature on use is limited. As mentioned previously, Moray et al. (2000) found that operators felt more self-confident if they performed the task manually before doing it with the automation than vice versa. Similarly, De Vries et al. (2003) found that the trust-self-confidence difference was lower when participants performed a task with automation first rather than completing it manually first. This likely suggests that, if self-confidence and trust are one of main factors driving automation use, doing a task with automation before completing it manually would lead to more automation use, but this is explicitly investigated here.

The current study aimed to understand the extent to which task difficulty, trust and self-confidence, and prior experience with automation influences the choice to use automation. To investigate these questions, a dual task paradigm was used in which participants completed one monitoring task (dispatching trucks as they filled with packages) and one visual search task (matching package barcodes). The monitoring task could be automated, such that the automation would visually alert the participants when the trucks needed dispatching, but the automation did not take the required steps to dispatch the truck.

It was first hypothesized that participants would choose to use automation more often in the difficult task than in the easy task condition, although the effect would be small, considering it was not captured in prior discretionary automation studies (H1). In examining order effects, it was hypothesized (H2) that participants who did the task with automation first would be more likely to use automation than participants who did the task manually first. Lastly, in line with previous research, it was hypothesized (H3) that when trust outweighs the participants' confidence in their manual skills, participants would use the automation more often, regardless of difficulty.

Methods

Participants

155 undergraduates participated for partial course credit. Three were removed due to clear inattention with zero responses to the main task. Two participants were removed due to issues saving data.

Task

Participants completed a dual task paradigm based on a warehouse management system similar to that in Barg-Walkow and Rogers (2016). On half of the screen, an image of a truck slowly filling with packages was present (see the left side of Figure 2). The truck could take anywhere from 19 seconds to 29 seconds to fill. Once it was full, participants had three seconds to click the "Dispatch" button, placed under the truck image, to remove the truck from the screen and make the next truck appear. Dispatching of the trucks was always done by the participant, but on half of the trials they were assisted by automation to alert them when the truck should be dispatched. The automation consisted of large red letters that appeared over the truck that said "DISPATCH". The automation was 86% reliable across all conditions. Out of 15 trials, it erred on two: one error was a miss (did not tell the operator to dispatch at all) and the other was a false alarm (told the operator to dispatch when the truck was not full). The false alarms always occurred when the truck was at least three-quarters full. The automation itself did not dispatch the truck, thus any missed dispatching or early dispatching were considered the fault of the operator.

The second task was a visual search task. On the other half of the screen, four letter strings (i.e. "OPBQ") appeared. One was the incoming package barcode and four others were matching barcode options (see right side of Figure 2). One of the four options matched the incoming package and participants used corresponding keys on the keyboard to indicate which package matched. They had four seconds to match the package before it disappeared and the next package appeared. The difficulty of this task was varied based on the theory of feature analysis in object recognition. In the easy task, the incoming package barcode was made of letters with dissimilar features to the matching packages. For example, the incoming package barcode could be “COPQ” and the options to match would be “XTKL”, “WTXV”, “COPQ” and “WNML”. In the difficult condition, the incoming package barcode was made of letters with similar features to the matching packages. For example, the incoming package barcode could be “OQPC” and the options to match would be “OQCU”, “PCUO”, “CUQU” and “OQPC”. Each of the matching options was labeled 1-4 and participants would press corresponding keys on the keyboard to indicate which matching option was correct.

Figure 2. A screenshot of a trial. The incoming package is LKKE and the correct package match is 2. The truck is not yet full. The grey area of the truck is the filled area, so once all boxes are grey, it would be full. The grey “Dispatch” button at the bottom would be used when it is full.

Design

All participants received practice blocks of the package and truck single tasks (with and without automation) and then practice trials of the dual task scenario (with and without automation). Practice blocks contained two truck examples or 10 package examples. Participants were then counterbalanced between the order of the easy and difficult conditions and whether they had the automation on or off in the first block. This created a mixed 2 (between: automation order) by 2 (within: difficulty) design. After experiencing both automation and no automation for each task difficulty, participants were given a choice to turn/keep the automation on or off. This was a binary choice and their decision was permanent for the following block (labeled "Choice" in Table 1). This created four conditions, shown in Table 2.

Table 2. All possible conditions in the current experiment. Participants were placed into one of the conditions. The presence of automation in the "Choice" block was based on the participant's decision.

	Block 1	Block 2	Choice	Block 3	Block 4	Choice
Condition	Easy Automation	Easy No Automation	Easy	Difficult Automation	Difficult No Automation	Difficult
	Easy No Automation	Easy Automation	Easy	Difficult No Automation	Difficult Automation	Difficult
	Difficult Automation	Difficult No Automation	Difficult	Easy Automation	Easy No Automation	Easy
	Difficult No Automation	Difficult Automation	Difficult	Easy No Automation	Easy Automation	Easy

Dependent Variables

The main variable of interest was the participant's decision to use the automation – their binary choice. Of additional interest was the accuracy of both the truck dispatching (did they dispatch the truck within the three second window) and the package task (how many packages did they correctly match). Reaction time for the package and truck tasks was also measured.

Additionally, the instantaneous self-assessment (ISA; Tattersall & Foord, 1996) was used to measure subjective workload after each non-automated block. Participants were provided with a chart to understand the 1-5 scale of the ISA and then responded on a slider. A four question trust scale from Lee and Moray (1992) was also completed on the automated system after each choice block, regardless of the choice. These questions were completed after the choice block so as not to prime the participants to consider their trust in the system more heavily than they would on their own when making the decision to use the automation. After completing the trust scales, participants were asked about their self-confidence in completing the truck task manually. Trust and self-confidence questions were responded to on a 1-10 scale. Specific scales can be accessed in Appendix A.

Results

Summary tables of the main results of this experiment are available in Appendix D.

Accuracy

To ensure the task difficulty (task load) manipulation was successful, the workload ratings and objective performance was analyzed. Performance was significantly worse for both tasks in the difficult condition than the easy condition. Package matching was 81% in the difficult condition and 93% in the easy condition ($t(142) = -24.29$, $p < .001$, $d = 2.10$). Truck dispatch accuracy was 57% in the difficult condition and 63% in the easy condition ($t(142) = -5.57$, $p < .001$, $d = 0.42$).

Subjective workload was also significantly different between conditions, with a higher rating (1-5 scale) in the difficult condition ($M=3.55$) than the easy condition ($M = 3.33$; $t(142) =$

-3.63, p < .001, d = .33). Therefore, the task difficulty conditions will now be referred to as "workload" conditions.

Automation Use

Participants were only given two alternatives to use the automation – yes or no – across two decision points and as a result, the data were non-normal. Therefore, non-parametric tests were used to analyze differences between groups as necessary. Overall automation use across all conditions was investigated using a Kruskal-Wallis test. There was no difference between the four conditions (K(3) = 4.27, p = .23). Therefore, the two conditions each under workload and order can be collapsed for further analyses.

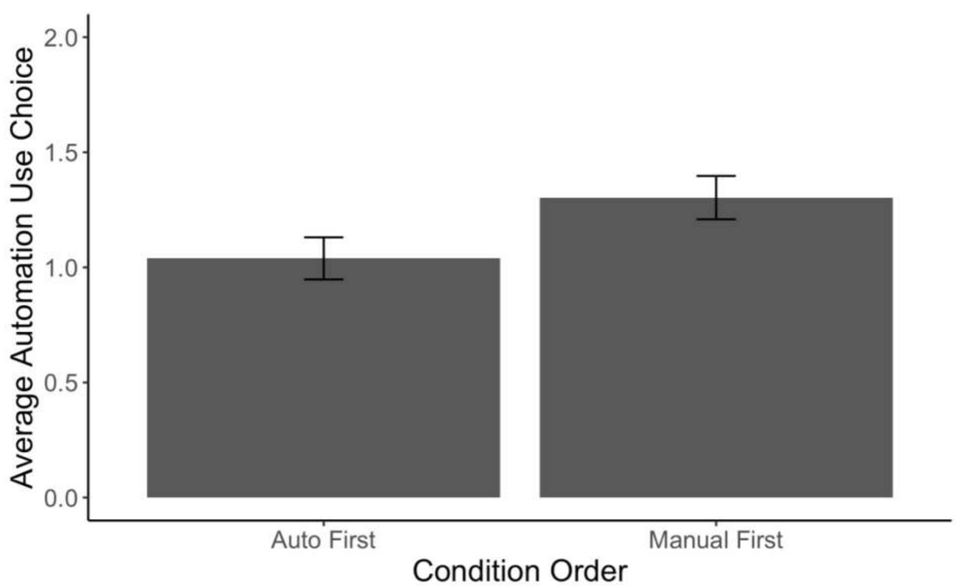

Figure 3. Average automation use (on a 0, 0.5 or 1 scale) by the order in which participants performed the task. Error bars represent one standard error of the mean.

Automation use was coded as a 0 if participants chose to do the task manually both times, 0.5 if they chose to do the task manually once and with automation once, and a 1 if they chose to do the task with automation both times. There was no difference in automation use for workload, irrespective of order. Using a paired Wilcox test, the difference between 60% and 55% of users turning automation on was not significant ($V = 739.5$, $p = .26$). This finding is in line with previous work on discretionary automation and suggests that, contrary to findings in static automation, the active choice to turn automation on or off is not a product of task difficulty or workload. However, when looking at the order effects, there was a significant difference (Figure 3). Participants in the manual-first condition used automation more often than those who were in the automation-first condition ($V = 2082$, $p = .04$). This suggests that learning to complete a task without automation increases the chances of using automation when given the choice – a contrast to the hypothesis that initially completing a task with automation would lead to increased use due to familiarity.

Trust and Self-Confidence

Overall, trust was not significantly different between workload conditions (M_{easy}: 5.49; $M_{difficult}$: 5.35; $t(142) = 1.81$, $p = .07$, $d = 0.08$), which is not surprising because the automation reliability was the same. Self-confidence was also not rated differently between the easy ($M = 6.55$) and difficult conditions ($M = 6.33$; $t(142) = 1.56$, $p = .11$, $d = 0.10$). This is in contrast to the objective performance metrics indicating worse performance in the more difficult condition.

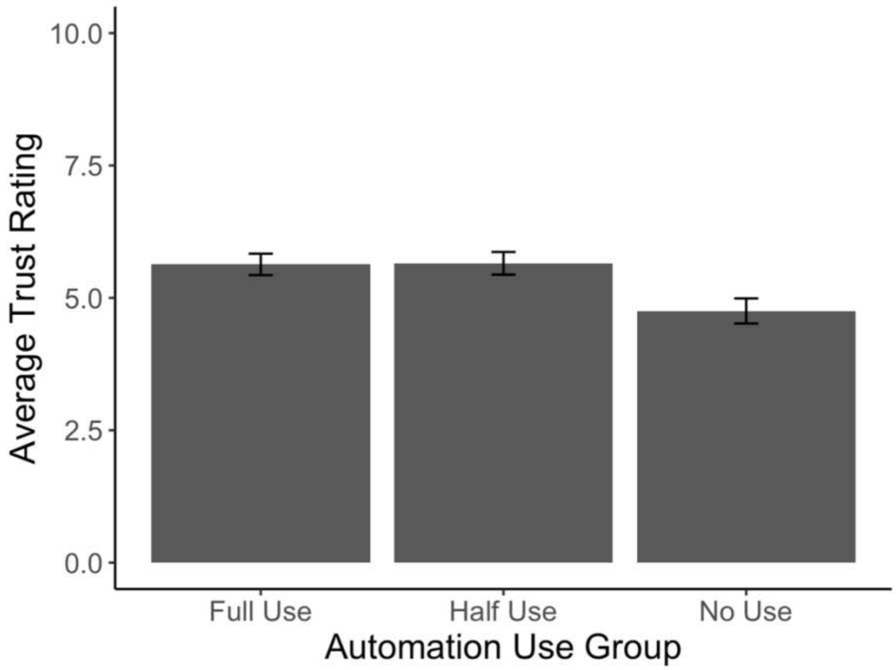

Figure 4. Average trust rating by the number of times automation was chosen to be used. Error bars represent one standard error of the mean.

To better investigate the influence of trust and self-confidence on automation use, participants were categorized as “no-use”, who never turned on the automation, “half-use”, who turned on the automation once, and “full-use”, who turned on the automation both times the choice was offered. Using a Krusal-Wallis test, subjective trust ratings were compared across user groups and showed a significant difference ($K(2) = 7.99$, $p = .01$; Figure 4). Follow-up Wilcox tests were conducted. No-use participants had significantly lower trust than participants in the half-use group ($W=593$, $p = .01$) and the full use condition ($W=695$, $p = .01$). There was no difference in trust between the half-use and full-use groups. This suggests that low trust may lead to no use, but higher levels of trust may not be as predictive of use.

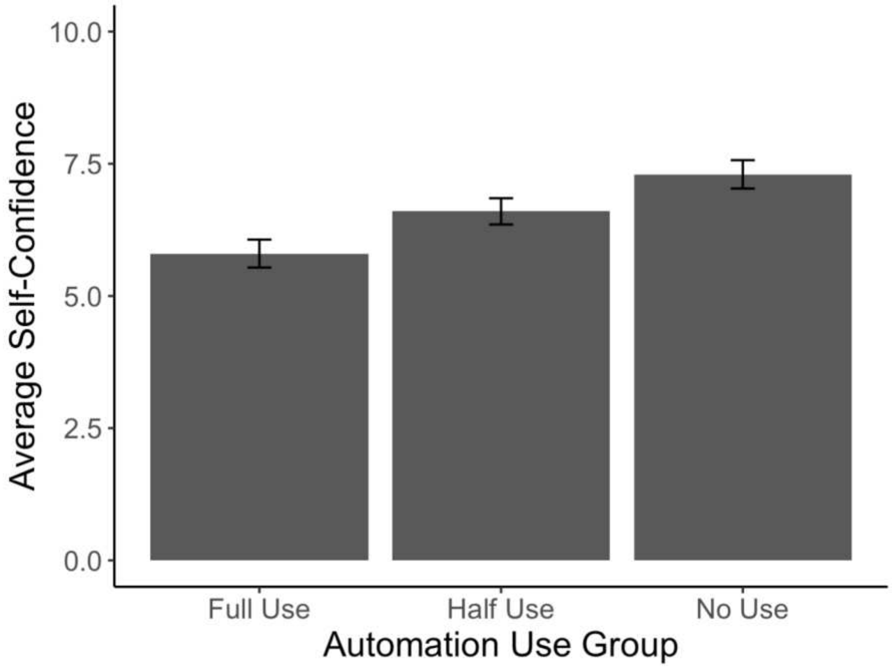

Figure 5. Average self-confidence ratings by the number of times automation was chosen to be used. Error bars represent one standard error of the mean.

Self-confidence was analyzed in the same way. A Krusal-Wallis test indicated differences between the groups ($K(2) = 13.87$, $p < .001$; Figure 5). Follow up Wilcox tests indicated that self-confidence was significantly higher in the no-use group than the half-use group ($W=1104$, $p =.03$) and the full-use group ($W=1463$, $p = .04$). The half-use group also had higher self-confidence than the full-use group ($W=1773$, $p = .04$). Therefore, as self-confidence rises, automation use decreases.

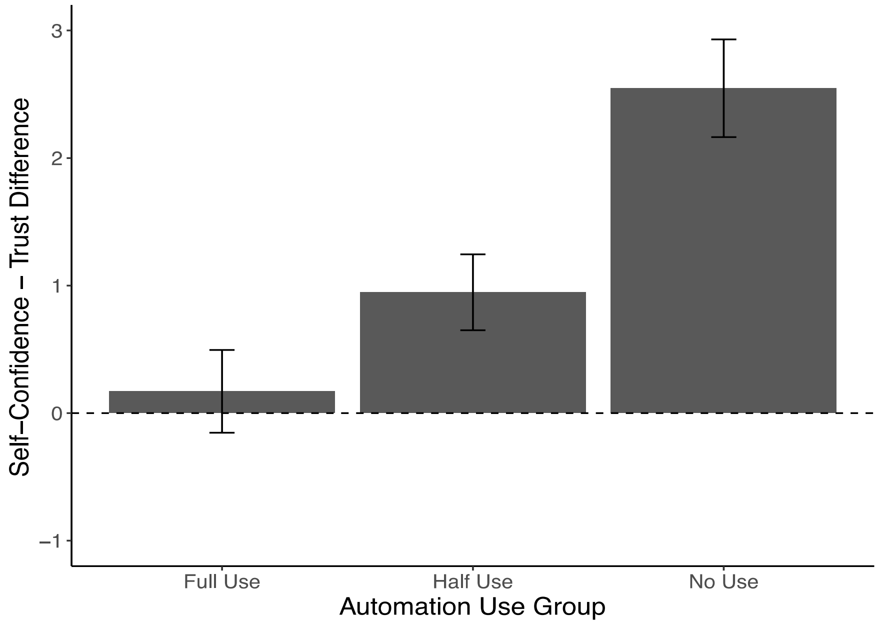

Figure 6. The difference between subjective self-confidence ratings and subjective trust ratings averaged by automation-use group. 0 on the y-axis indicates the same level of trust and self-confidence. Taller bars represent more self-confidence than trust. Error bars represent one standard error of the mean.

The difference between self-confidence and trust was calculated for each participant and then averaged by use groups (Figure 6). Again here, a Krusal-Wallis test indicated differences between the groups ($K(2) = 19.33$, $p < .001$). Follow-up Wilcox tests showed a much larger difference for the no-use group (far more self-confidence than trust) compared to the half-use group ($W=537$, $p < .01$) and the full-use group ($W=491.5$, $p < .001$). The half-use group also had a larger average difference than the full-use group ($W=1141$, $p = .05$). The large differences between the groups indicates that the self-confidence-trust difference is a driving force behind the choice to use automation.

The difference in trust and self-confidence was then analyzed as a function of order. There was no significant difference between the manual-first ($M=1.00$) and automation-first

(M=1.03) groups (W=2557.5, p =.94). This suggests that the order effects are not driven by trust and self-confidence.

Discussion

The current experiment aimed to investigate the impact of three main variables on automation use: workload, order of experience (of automation), and the trust-self-confidence relationship. In line with previous literature (Riley, 1995) but against the first hypothesis, there was no effect of workload of the concurrent task on the choice to use automation. This experiment explicitly measured workload subjectively so that it could be differentiated from task difficulty, although the two were related in this context. This finding is in contrast to the literature on static automation, where people are generally more likely to rely on automation when the task is more difficult (Dixon et al., 2007). The reason for this disconnect is unclear, however, the self-confidence ratings provide some insight. Self-confidence subjective reports were not different between the easy and difficult condition, which indicates that although participants objectively performed worse and subjectively rated workload higher in the difficult condition, they did not think they were any worse at the task when it was difficult. Other literature has found overestimations in one's task performance, especially for difficult tasks over easy tasks (Larrick et al., 2007; Merkle & Weber, 2011; Moore & Healy, 2008). If participants felt they could perform equally as well in the two conditions, their perceptions of performance may have provided an (incorrect) indication that they did not need the automation's assistance.

Prior experience did drive automation use, to an extent. Contrary to the second hypothesis, participants who performed the task manually first (and then experienced automation before making their choice) were more likely to use automation than those who experienced the opposite order. This was surprising, as intuitively we see people who experience automation or learn a task with automation often more likely to use it – consider young children's comfort with smart phones compared to their grandparents' use. However, this is in line with the results of De Vries et al. (2003), who found that participants who did the task with automation first had closer levels of trust and self-confidence, which would suggest less automation use due to a bias to rely on manual efforts. Here, however, analyses showed that there was no difference in trust and self-confidence between the two order groups, so this is not what is driving the automation use difference. Instead, it is possible that by experiencing the task manually first, the introduction of automation made the benefit of the automation more salient, leading those participants to use it more. More research is needed to investigate this effect.

Lastly, in line with the findings of Lee and Moray (1992) and Riley (1995), trust and self-confidence were clearly related to automation use. Trust in the automation alone was lower for participants who did not use the automation at all, but higher trust was not indicative of more use. This provides some evidence for the third hypothesis. Self-confidence, however, was inversely related to automation use on all levels. The most striking differences appear when trust and self-confidence were considered in relation to one another. Participants with high trust and low self-confidence were most likely to use the automation, confirming the second half of the third hypothesis. As trust and self-confidence became more similar, automation use decreased. This was true regardless of order effects or workload condition, suggesting that it is one of the main drivers of automation use in this context.

Conclusions

The current experiment found strong evidence for the impact of trust and self-confidence on automation use. While the finding itself is not novel, the proof of concept in a new paradigm, with a large sample size, and fewer manipulated variables than previous findings suggests that the relationship between trust and self-confidence is a consistent and strong driver of automation use. This experiment also provided additional evidence against both subjective and objective workload impacting automation use. This null effect is lacking a clear explanation, especially considering the stark contrast to static automation dependence literature (Biros et al., 2004; Dixon et al., 2007; McBride et al., 2011; Xu et al., 2007). However, when considering the effect of self-confidence and trust on automation use, and the lack of a difference in self-confidence between difficulty conditions, the results of this experiment suggest that the choice to use automation may be driven by operator perceptions of performance, rather than objective metrics. A first step in understanding these effects is to ensure they replicate in other paradigms, and also to make the difference between workload conditions more salient.

These findings are important for designers and real world implications. If workload or difficulty is not a driving factor of automation use, designers should not assume that introducing reliable discretionary automation into a difficult task will improve performance. Automation is often implemented in scenarios where humans may struggle to perform at a level acceptable for the specific context. If operators are not sensitive to the need for automation when concurrent tasks are more difficult, allowing them to control when the automation is present may result in disuse or misuse, which would lead to less than optimal performance.

This experiment also hints at the potential importance for operator perceptions. The lack of a change in automation use paired with the lack of change in self-confidence indicates the importance of this variable that is seldom considered in static automation. This provides a first hint that while objective metrics may drive automation *dependence*, other factors, specifically perceptual metrics, may drive the *choice* to use automation.

CHAPTER 4 - EXPERIMENT 2 (Task Load Impacts on Automation Use)

Introduction

Experiment 1 indicated that subjective workload or task difficulty (task load) of a concurrent task does not exert an influence on the choice to use automation. However, it is possible that the difficulty difference between the “easy” and “difficult” conditions was not salient enough to participants to influence automation use, which is indicated by the lack of a difference in self-confidence ratings across conditions. Additionally, the difficulty was varied on the concurrent task (which was not automated), which could cause participants to view automation use as unrelated to task difficulty. Therefore, the current experiment examined how a range of difficulties (task loads) of the automated task itself influenced automation use. Using a single task paradigm, automation was presented across four difficulty conditions and was used as an aid for detecting hostile ships through movement. It was hypothesized that automation use will increase as workload increases, but the effect will be small.

This experiment also investigated participant accuracy as a factor involved in automation use. As mentioned previously, this area of research is sparse. Riley (1995) found that manual performance was not indicative of automation use. While replicating this effect is of interest, this pathway also allows us to investigate other aspects of operator accuracy. It has been hypothesized that the extent to which automation improves performance may impact automation use (Hoff & Bashir, 2015), but has not been explicitly investigated. Here, it is hypothesized that participants whose unaided performance is improved most by the presence of automation will use automation more often.

Methods

Participants

51 people participated in the experiment through online data collection on Prolific. One dataset was removed due to using exactly the minimum number of steps on each trial and performing at 0% accuracy.

Task

Participants viewed a computer screen containing a large circle labeled “you”, indicating their ship’s position, which they could control, and circles with numbers which represented other ships and were controlled by a software application as shown in Figure 7. Trials began with the starting location of all ships randomly computed. The participants could move their ship by clicking direction keys on the screen, but an arrow key could only be clicked once per second to negate the potential to create apparent motion through rapid keystrokes. The user ship could move up, down, left, or right. Each time the user ship moved, the computer-controlled ships also moved, although these ships were able to move diagonally. Thus, all ships moved at the same time, with at least a 1s delay between movements.

On each trial, one of the computer-controlled ship’s movements was dependent on the user’s movements in a pattern described below, and the other five moved independently of the user. Two trials were given as practice; therefore, no data were recorded. The practice trials involved one hunting behavior and one shadowing behavior with a 25% chance of random movement. On each practice trial, the hostile ship was a different color and the hostile behavior was announced before the trial started. This allowed participants to practice working through a scenario but also showed the difference between hostile behaviors.

The hostile ship would do one of two things—hunt or shadow. Hunting meant selecting the movement to best eventually reach the user ship. An algorithm computed which directional movement produced the greatest reduction in distance between the two and moved the ship in that direction as the user ship moved. Shadowing aimed to generally keep a consistent distance from the user ship through replication of their movements. For instance, if the user moved left, the shadowing ship also moved left. If the user ship moved toward the shadowing ship, it moved the same direction as the user so the distance between the ships stayed the same. These target movements occurred simultaneously with the user ship movement that triggered it.

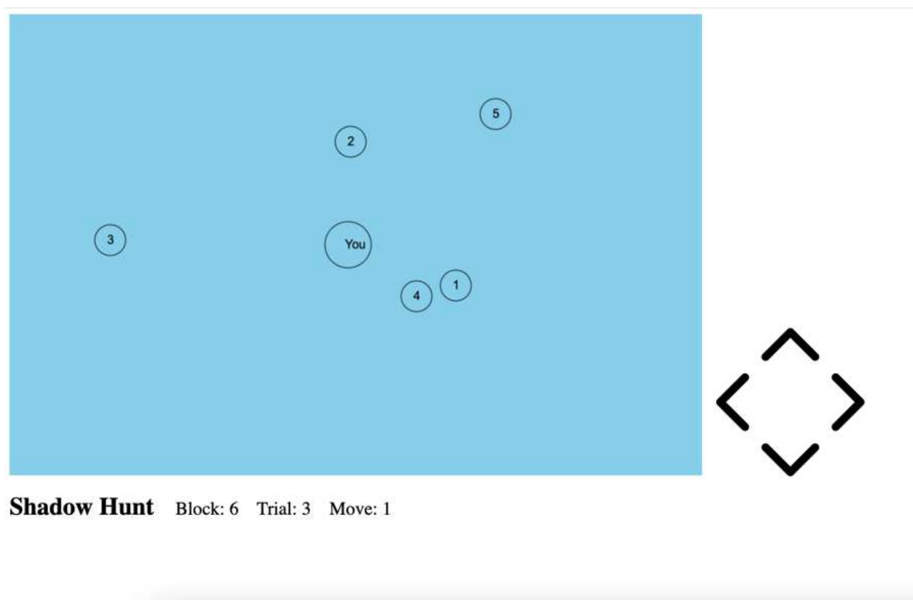

Figure 7. Example of a 5-ship trial. The large circle is the usership and the numbered circles represent other ships in the area, one of which is hostile.

The non-hostile ships were randomly assigned one of two other movement patterns. They could have exhibited "patrol" behaviors, where they moved in a rectangular course around the screen. The rectangular path could be oriented in any direction and the ship could start at any point on the path. All movements of the distractor ships contained the same amount of noise as the hostile ship so that targets could not be distinguished based on differing randomness of the path followed. The other non-hostile behavior option was “distractor”, which is when the ship moved toward the starting point of the user ship. The goal destination did not change as the user ship moved.

On each trial, the participant was required to make at least six moves, but no more than 35 moves, in any pattern they chose before determining which ship they believed was hostile. Once they made a decision, they clicked on the “End” button, which introduced a question across the top of the screen asking the participant to select if they were being hunted, shadowed, or neither. If they chose hunting or shadowing, the next question asked them to choose which ship was exhibiting that behavior by clicking the radio button that matched the ship number they believed was hostile. They then clicked “submit” and were given feedback only on the correctness of their response, but not on the correct target nor the hostile behavior exhibited on the trial.

Automated Aid

One machine learning model was trained on a dataset of shadowing ships and another was trained on a dataset of hunting ships. Both models were trained on datasets that included 26 distinct features for the models to utilize when determining the hostile intentions of a ship. For every move after the 5^{th} move, both machine learning models classified each ship into two categories, hostile and non-hostile. If the model trained on the shadowing dataset returned any ships classified as hostile, the hostile-classified ships from the shadowing model would be displayed. If it did not, the hostile-classified ships from the hunting model would be displayed. This information was displayed by highlighting the ships outer edge either yellow or red based on the confidence/probability of the classification returned by the machine learning model. The threshold of which color was displayed was calculated using Bayesian Smoothing, regardless of which model returned the hostile classification. If the returned value of the Bayesian Smoothing equation was greater than or equal to 0.75 normalized (75% confidence), the ship would be highlighted red. If the returned value of the Bayesian Smoothing equation was greater than or equal to 0.5 normalized (50% confidence), the ship would be highlighted yellow. This resulted in varying numbers of ships and colors presented on each movement in a trial. If the aid did not detect any ship as potentially hostile with 50% or more confidence, no ship was highlighted. The aid was correctly highlighting the hostile ship 74% of the time at the end of a trial on average, ranging from 89% with 5 ships to 62% with twenty ships. This includes trials where more than one ship was highlighted.

Design

Participants first completed two practice trials, one of shadowing and one of hunting movements. Participants were informed which ship was exhibiting which behavior so they could learn how to complete the task. Then, they completed 10 blocks of four trials. Each of the four trials would have a task load of 5, 10, 15 or 20 ships in addition to the user ship. They were presented in a random order within each block. Regardless of the number of ships, there was only one hostile ship and it only exhibited hunting or shadowing behavior. The number of ships was the manipulation of difficulty. Previous research indicates that increasing the ship numbers lowers performance (Patton et al., 2022).

The first four blocks were performed manually. The next four blocks were performed with the automated aid. In the last two blocks, participants were given a choice to turn the automation on, or (the default) leave it off. Participants were able to make their decision regarding the automation up until the end of move 6 on each trial. The automation toggle froze for the remainder of the trial and reset at the start of each new trial. The automation operated the same in the choice blocks as in the automated blocks earlier in the experiment.

Results

Summary tables of main results are available in Appendix D.

Accuracy

To ensure that the aid was beneficial and that more ships made the task harder, accuracy across conditions was checked. Accurate detection means that the participant correctly identified both the correct ship and behavior, not one or the other. Overall detection accuracy was higher with the aid (43%) than without (26%; $F(1,49) = 23.22$, $p < .001$, $\eta_p^2 = .32$). Detection accuracy also differed as a function of task load ($F(3, 147) = 22.44$, $p < .001$, $\eta_p^2 = .31$; see Appendix D). There was no significant interaction between ship number and aid presence ($F(3,147) = 1.24$, $p = .29$, $\eta_p^2 < .01$). In the interest of investigating the most salient task difficulty difference, accuracy and automation use between five and 20 ships was analyzed. There was a large difference in overall accuracy between five ships (M=46%; M = 38% above chance) and 20 ships (M=24%, M = 24% above chance; $t(49) = 8.03$, $p < .001$, $d = 0.96$), indicating that the task was more difficult at 20 ships.

Automation Use

Overall, participants turned on the automation and left it on 67% of the time. This was an active choice, as the default was set to "off" in the choice trials. Participants who turned the automation on performed better (M=47%) than those who left it off (M=19%; $t(77) = 2.68$, $p = .008$, $d = 0.60$).

Once again, overall task difficulty did not play a role in automation use (Figure 8), with no difference across the four levels of difficulty (numbers of ships) ($F(3,147) = 1.90$, $p = .13$,

η_p^2= .03). In the interest of the most salient difference in task difficulty, a follow-up t-test was conducted on automation use rates between five and 20 ships. Automation use was significantly less in the five ship condition (M=52%) than the 20 ship condition (M=63%; $t(49) = -2.15$, $p = .03$, $d = 0.26$). This indicates that difficulty of the automated task may influence the choice to use automation, when the differences in difficulty are large.

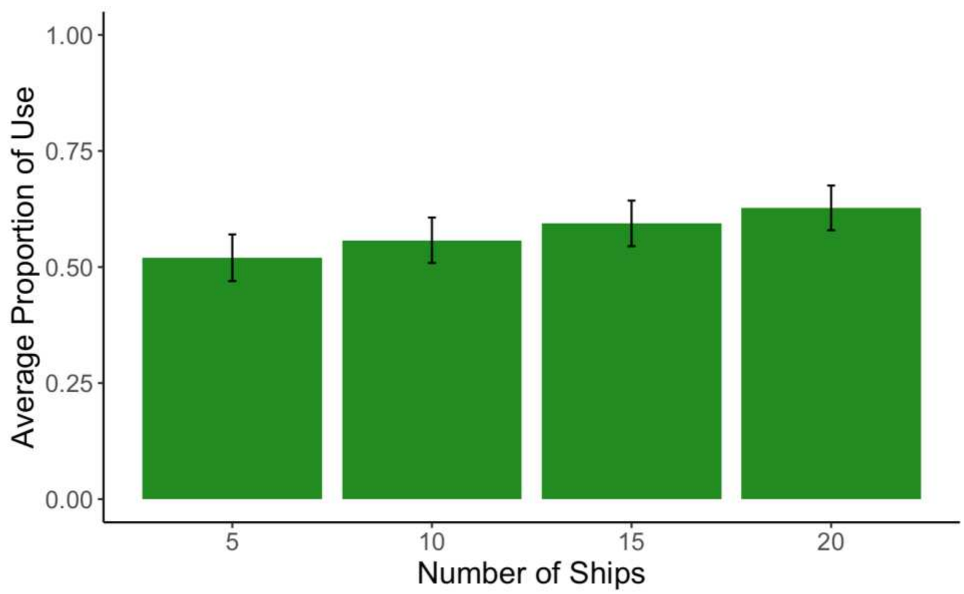

Figure 8. Average use of automation as a proportion by the number of ships displayed. Error bars represent one standard error of the mean.

Also of interest in this experiment was whether or not individual differences in participant accuracy would impact their use of the automation. First, a correlation of participant unaided accuracy with automation use showed no evidence of a significant relationship (r=.007, p = .96), indicating that poor manual performance on this task was not related to more automation use. Then, the difference between unaided and aided accuracy (aided minus unaided) for each participant was calculated and correlated with their automation use. The correlation (r=.20, p = .06) just failed to reach significance but hinted at the potential for small positive relationship between the two variables. When (13) participants who performed worse with the automation were removed, the relationship was smaller and no longer approached significance ($r=.22$, $p=.19$), indicating that it is unlikely the improvement in performance with the aid influences the decision to use automation.

Trust and Self-Confidence

As in Experiment 1, trust and self-confidence were examined here. Due to the large variation in individual automation use because of having eight trials with automation choices, correlations were used rather than binning use into categories as in Experiment 1. Trust was slightly, but not significantly, related to automation use ($r=.25$, $p = .08$). The difference between trust and self-confidence was more similarly related to automation use ($r=.27$, $p = .06$), although there were participants with high trust and low self-confidence that never used the automation, and vice versa.

Discussion

This experiment aimed to better understand the influence (or lack thereof) of task difficulty on automation use and replicate the findings of the trust-self-confidence relationship on automation use. As predicted, the larger the difference between trust and self-confidence, the more automation was turned on, regardless of task difficulty or other factors. This is in line with Experiment 1 and previous research (Riley, 1995; Lee & Moray, 1992).

The findings on task difficulty are slightly more complex. Overall, a one-way ANOVA indicated no difference in automation use across the various ship numbers (as the task difficulty changed). However, it was hypothesized that task difficulty may not be driving automation use in Experiment 1 or in Riley's (1995) previous experiments because the difference was not salient enough or, in the case of Experiment, because it was the concurrent task and not the automated task where workload was varied. Therefore, the automation use and accuracy at five ships and 20 ships was tested against each other. Overall accuracy was much higher and automation use slightly lower in the five ship condition compared to the 20 ship condition, supporting the first hypothesis. This indicates that task difficulty may actually play a role in the choice to use automation, although the effect is small. Post-hoc, the null results of task difficulty in Experiment 1 were thought to be due to participants not realizing the impact the more difficult task had on their performance, as their self-confidence ratings did not change. Here, it seems that the difference in difficulty between five and twenty ships may have been salient enough to nudge participants toward automation use. Self-confidence was only collected at the end of this experiment, which limits the ability to investigate how it changed across difficulties. It also does not provide evidence for whether it is difficulty alone, or a change in self-confidence due to difficulty, that is driving automation use. This is investigated in Experiment 3.

The other factor investigated in this experiment was individual differences in participant accuracy. Correlations indicated no relationship between unaided participant accuracy and automation use, nor was there much of a relationship between automation use and the difference between aided and unaided participant accuracy ($p=.06$), which opposes the second hypothesis that participants whose unaided performance was improved most by the presence of automation would use automation more often. This is surprising, as participants who performed poorly would have been expected to use automation more often, especially if it was substantially improving performance. However, just as before, it may be that participants did not understand just how much the automation was improving their performance and therefore could not accurately calculate the cost-benefit trade-off for automation use (Hoff & Bashir, 2015).

Although participants were told after each trial whether their answer was correct or incorrect and their score shown across trials, previous research suggests that this task is difficult enough that additional information can overburden participants and they may not pay attention to it or be able to properly process it (Patton et al., *under review*). Participants were also not given feedback about whether they or the automation was wrong, which could make it harder to see the benefit of the automation.

These findings suggest that humans are not prone to making the optimal decision when choosing whether to use automation or not. Some participants continued to choose not to use automation, even when overall accuracy without automation was below 20% - a clear non-optimal choice. While humans are known to be imperfect decision makers (i.e, Klein, 2008), in this context it becomes crucial to understand why they are making these decisions. Experiment 1 hinted at the importance of perception of workload in the choice to use automation, and that may still be true. The current experiment examined objective metrics surrounding automation use – specifically objective task load and changes in performance accuracy – but lacked the ability to investigate perceptions of self-confidence and subjective difficulty between the four task load conditions or to understand people's perceptions of the usefulness of the automation. Understanding why the influence of difficulty is unstable becomes important as real world applications for this work are considered.

Conclusions

This experiment suggests that if changes to the difficulty of a task that can be automated are salient, it may influence automation use. Additionally, the difference between trust and self-confidence continues to be a clearly related to automation use, as increasing trust and decreasing self-confidence leads to more automation use. Surprisingly, the connection between operator accuracy – both aided and when calculated as the difference between aided and unaided – was not a clear indicator of automation use. This all links back to the idea that it may be subjective, not objective, perceptions of variables that influence automation use. Trust and self-confidence are subjective measures and seem to consistently be related to automation use. When objective measures are used, their predictive value for automation use is not as strong as the subjective measures.

These findings also suggest that real world automation use may not be optimal, especially under difficult task loads. Although automation is often introduced into tasks that are difficult for humans, the improvement to performance that it provides may not be enough to convince operators to use it, especially if they are highly confident in their abilities. Rather, automation can be expected to be used when self-confidence in completing a task is low. However, it is still unclear how task difficulty may play into the choice to use automation – a crucial piece of the story for real world applications.

CHAPTER 5 - EXPERIMENT 3 (Subjective Versus Objective Workload on Automation Use)

Introduction

The results of Experiments 1 and 2, in conjunction with previous research, strongly suggest that trust and self-confidence together are predictors of automation use. However, the findings on task difficulty are mixed. The missing piece may in perception. As seen in Experiment 1, although the concurrent task was objectively more difficult, subjective self-confidence did not change and neither did automation use on the primary task. In Experiment 2, task difficulty of the primary automated task did change automation use at the most extreme ends, but self-confidence and subjective workload were not recorded across the different difficulty conditions, so whether self-confidence was impacted was unknown.

It seems that a consistency across these data is the difference in subjective versus objective perceptions of the task and its associated factors. Trust and self-confidence are subjective, and clearly related to automation use. Operator accuracy and task difficulty, thus far, have been objective – such that they are measured through performance and task load rather than questionnaires for the participants – and they are less clearly related to automation use than self-confidence and trust. It seems then, that subjective perceptions may play a role. Thus, it may be that if a task is perceived to be difficult, independently of its objective difficulty, a person is more likely to use automation. One study has looked at a tangential issue. Schwark et al., (2010) found that manipulating perceptions of trial difficulty – telling participants the trial would be difficult regardless of objective difficulty – was enough to increase agreement rates with an automated system. While this study focused on a static automation system and failed to assess participants' perceptions of the difficulty, it does provide a foundation to suggest that perception is important.

There is evidence that perceptive and objective measures of tasks or preferences differ from each other. For example, Horrey and colleagues (2009) found that drivers rated a dual driving task as no more difficult than a single driving task, even though objective performance significantly decreased. We also see this in the performance-preference dissociation, where users want what is not best for them (Forster et al., 2019; Ziv & Lidor, 2021). It would therefore not be surprising if participants did not accurately perceive the difficulty of the tasks in the previous experiments, leading to nonoptimal automation use.

Therefore, the current experiment aimed to understand if it is perceptions of task difficulty, objective task difficulty, or both, that drive automation use. To do this, the goal was to change perceptions of difficulty without changing objective difficulty. This way, the two can be separated. Intuitively, one way to do this is to change the reference point of the participant. We see this in visual perception, when a grey square looks darker next to a white square but lighter next to a black square, even if the grey color does not change (Gilchrist et al., 1999). There is some evidence that this works for cognitive decisions and perceptions as well. The anchoring bias (Tversky & Kahneman, 1974) is one way that a reference point can change perception. This is seen in the automation literature in a study done by Barg-Walkow and Rogers (2016), who provided participants with text instructions about the reliability of an automated system. The provided reliability information was either higher or lower than the actual reliability. Once participants interacted with the system, their estimated reliability moved closer to the real reliability, but did not reach it, as a result of the initial anchor.

This appears to be true in difficulty ratings as well. There is evidence that participants are sensitive to difficulty changes (i.e., Robinson, 2001) and it seems that participants will use what are termed "central cues" to determine difficulty (Vangsness & Young, 2019). This means they use information available to them at task onset that is independent of skill and resource allocation to estimate how difficult a task will be. For example, seeing an increase in objects on a screen is likely to make a participant think the current task will be more difficult than the one before it. One study investigated a tangential issue, where they told participants about either the easy or difficult option of a task, where difficulty was manipulated by the number of items that had to be visually found (Voslinsky & Azar, 2022). Participants then received either the easy or the difficult version, and their affect was measured. Those that received the difficult task after being told about the easy task had stronger negative affect, and those that received the easy task after being told about the difficult task had more positive affect. Although perceived difficulty was not measured, it seems that the initial anchor had an impact on the perception of the task.

Measurement of Subjective Variables

Of additional interest in this study was the measurement of the subjective variables trust and self-confidence. Although these two variables, and the difference between them, have been used in prior literature (i.e., Riley, 1995; Lee & Moray, 1992), some literature suggests that "difference scores", or the subtraction of self-confidence from trust to create a new metric, are problematic (Cronbach, 1958; Edwards, 1994). Specifically, one prevalent thought is that if a difference score does exhibit reliability, then the components of the difference score should be more reliable (Edwards, 1994). This is because the noise in the data can be amplified when a difference score is used, because the variance of a difference is proportional to the sum of the variance of its components, hence creating a "nosier" and thus less reliable measure. Therefore, a new combined measure of trust and self-confidence was created in an attempt to remove the difference score problem. Participants were asked how much they agreed with the statements “The automation is better than me at this task” and “I am better than the automation at this task”. The second question was then reverse coded and the two responses averaged. This provided one metric for a less noisy indicator of the trust-self-confidence difference. This measure will then be compared to the difference score to compare their abilities to predict automation use.

Additionally, the variable of self-confidence is measured through a single question without much prior research on its validity or reliability. Self-efficacy, on the other hand, has research surrounding its validity (Bandura, 2006; Burrell et al., 2018). Self-efficacy is defined as the “beliefs in one’s capabilities to organize and execute the course of action required to produce given attainments” (Bandura, 1977). On the surface, it does not seem very different from self-confidence, but generally is thought to be more context specific. Combined with its extensive use in research and available information on validity, it may be a better measure than a simple self-confidence rating. This is investigated in the current study, tangential to the main question of automation use.

Pilot Study

To ensure that the various conditions are perceived as differentially difficult, a small pilot study was conducted. Using the base paradigm from Experiment 2, participants completed trials with either two (low anchor) and then 15 ships, or 28 (high anchor) and then 15 ships. They were then given the choice to use automation for both sets of ship numbers, but the interest was in the automation use between groups for 15 ships. The goal was to see if the different anchor difficulties (two versus 28 ships) changed the perceived difficulty of 15 ships. There were 14 people in the high anchor condition and 16 people in the low anchor condition. They completed the entire study, and their difficulty ratings and on workload ratings on the 15 ship manual trials were compared. Results indicated that, on the ISA (workload measure), the difficult anchor group actually rated the 15 ship condition slightly more difficult ($M=3.57$, $SD = 0.85$) than the easy anchor group did ($M = 3.0$, $SD = 0.81$; $t(28) = -1.87$, $p = .07$, $d = 0.68$). This was trending in the opposite direction as expected, but would still allow for a comparison of perceived difficulty differences. Although the p-value was not quite significant, power was low with such a small sample size, and so the full sample was collected with the current methodology.

Current Study

The current experiment followed the same methodology as the pilot study. Participants were assigned to the low or high anchor group, and after each block of trials they reported their perceptions of difficulty, self-confidence, trust in the automation, along with their self-efficacy and responded to the new trust-self-confidence measures. It was hypothesized that:

1. Trust and self-confidence, both individually and together through the "self-confidence minus trust" metric, will again be predictors of automation use.

2a. Automation use will to the extent that perceptions of workload differ, such that a task rated as more difficult will result in more automation use than a task rated as less difficult.

2b. Specifically, automation use will be higher in the 15 ship condition for participants whose reference point was two ships, compared to the participants whose reference point was 28 ships because they will perceive the 15 ships as being more difficult when it followed two ships.

Methods

Participants

An a-priori power analysis for a t-test to determine the difference between the automation use in the two reference point conditions suggests 88 people total to detect an effect size of .5 with 95% power. 90 total participants were collected from Prolific, with 45 in each condition. Four participants from the low-anchor condition were removed due to overall accuracy below 25% and an average number of steps below 7. Low accuracy, even with the automated aid, and an average number of steps near minimum indicates a lack of effort from the participant. Additionally, the most important analyses were conducted with and without these participants and the significance level did not change.

Task and Design

The task from Experiment 2 (detecting hostile ships) was used as the base paradigm. Participants were assigned to one of two groups: the “easy” anchor or the “difficult” anchor. The easy anchor was a block of trials with two ships, and the difficult anchor was a block of trials with 28 ships. All participants also saw trials with 15 ships. The goal was to change the perceived difficulty of 15 ships based on the anchor – it would be seen as more difficult in the easy anchor condition and less difficult in the hard anchor condition.

All participants performed blocks of five trials per number of ships. These trials were blocked together, for example, participants saw five trials with two ships and then five trials with fifteen ships. Both groups saw the 15 ship block second, so as to provide the other ship number as a stronger anchor of difficulty (see Figure 9). Participants completed both blocks manually and with automation (the same automated highlighting aid as Experiment 2). After completing the task manually and with automation, participants again completed both blocks but this time, they were given the choice to have automation present or absent on each trial.

After completing the choice block, participants completed two blocks of trials with the number of ships they did not previously see (two or 28), one with and one without automation. Then then completed a block with 15 ships and the choice to use automation. This provides a different reference point for task difficulty of 15 ships and should bring the perceived difficulty back to a similar level between groups.

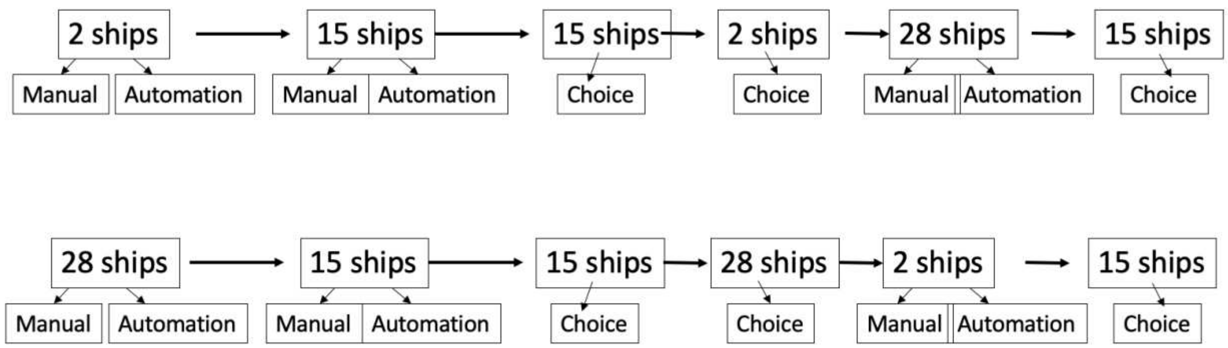

Figure 9. Examples of the flow of ship number trials for the two groups of participants. Note the first two blocks will be counterbalanced, as will the third and fourth blocks. These diagrams do not include the perception questions that occur between each block.

After every block, participants were asked to rate their workload in the ISA. They also were asked to rate how difficult the task was on a scale of 1-5. After the blocks where they had a choice to use automation, participants were asked the workload and difficulty questions, as well as questions on trust and self-confidence, as in Experiment 2. For the current experiment, Cronbach's alpha was calculated for the trust questionnaire, and indicated excellent internal consistency ($\alpha = 0.87$, 95% CI [0.83,0.90]). Additionally, a question on self-efficacy, adapted from Bandura's guide for constructing scales, was added to compare to self-confidence. Participants also rated their agreement with two statements about their abilities versus the automation's abilities in this task - "The automation is better than me at this task" and "I am better than the automation at this task". This allowed for a more direct way of comparing trust and self-confidence.

Results

Summary tables of main results are available in Appendix D.

Accuracy

Accuracy was measured as the correct identification of the hostile ship number and behavior. As expected, performance differed across the three difficulty conditions when averaged across all blocks that contained each number of ships (Row 1 in Table 2; $F(2,170) = 96.32$, $p < .001$, $\eta_p^2 = 0.46$). Performance was quite high with two ships (M=74%) but dropped off quickly at 15 ships (M=56%), which was a significant difference ($t(86) = 10.57$, $p < .001$, $d = 0.93$). Performance continued to drop from 15 to 28 ships (M=49%; $t(86) = 3.96$, $p < .001$, $d = 0.31$).

Due to a difference in the level of chance based on the number of ships, accuracy was re-calculated and analyzed. The chance of randomly choosing the correct ship number was multiplied by the chance of randomly choosing the correct behavior, and chance level for 2, 15 and 28 ships was 25%, 3% and 1%, respectively. A repeated measures ANOVA indicated the potential for differences between groups ($F(2,170) = 2.76$, $p = .06$, $\eta_p^2 = .03$). Follow up t-tests indicated that the change in accuracy above chance from 15 ships (52%) to 28 ships (48%) was significant ($t(85) = 2.71$, $p = .008$, $d = 0.22$). The change from 2 ships (50%) to 15 ships was not significant ($t(85) = -1.32$, $p = .18$, $d = 0.12$). The difference was 2 to 28 ships was also not significant ($t(85) = 0.93$, $p = .35$, $d = 0.10$).

Perceptions of Difficulty

As one goal of the study was to change perceptions of difficulty of 15 ships, the difficulty ratings between high and low anchor groups for 15 ships without automation were evaluated (Table 3). An independent samples t-test indicated that there was not a significant difference in perceived difficulty between the low anchor (M=3.97, SD=0.98) and high anchor groups (M = 4.15, SD=0.85; $t(84) = -0.90$, $p = .36$, d= .19). This difference was even smaller when compared for the ISA ratings ($t(84) = -0.07$, $p = 0.93$, $d = 0.01$). This indicates that the manipulation for changing perceived difficulty based on an anchor was not successful.

Overall perceptions of difficulty did, however, match objective difficulty (performance) metrics. As shown in Table 3, collapsed across the two groups for the manual (no automation) trials, participants rated the two ship condition easier (M = 2.37) than the 15 ship condition (M = 4.06; $t(85) = 12.72$, $p < .001$, $d = 1.68$). Participants also rated the 15 ship condition easier than the 28 ship condition (M = 4.29; $t(85) = -2.31$, $p = .02$, $d = 0.23$), although the effect is small, just as it was for performance accuracy.

Table 3. Averages across groups for 2, 15 and 28 ships. Significant differences were seen across all conditions for accuracy, and only between 15 and 28 ships for accuracy above chance. No differences were seen between 15 and 28 ships for perceived difficulty, automation use, or the ISA.

	2 Ships	15 Ships	28 Ships
Accuracy	74%	56%	49%
Accuracy Above Chance	50%	52%	48%
Perceived Difficulty	2.37	4.06	4.29
Automation Use	45%	78%	83%
Trust Minus Self-Confidence	-0.11	0.97	1.18

ISA	2.59	3.15	3.23

Table 4. Data for 15 ships as a function of prior shipload.

	2 - 15 ships	28 - 15 ships
Perceived Difficulty	3.97	4.15
ISA	3.34	3.35
Performance	53%	58%
Automation Use	70%	83%

Automation Use

Between conditions, automation use for the initial block of 15 ships was compared (Table 4). Participants in the low anchor group used automation 73% of the time, and participants in the high anchor group used automation 80% of the time, but this difference was not significant ($t(84) = -1.28$, $p = .20$, $d = 0.27$). This provides disconfirming evidence for the second hypothesis, that automation use would be higher in the 2 ship anchor group as a result of increased difficulty perceptions. However, this is in line with the lack of a difference in perceived difficulty between the two 15 ship conditions.

Automation use was investigated across the difficulty conditions to see if task load (measured by performance) played a role in the choice to use automation (Figure 10). Collapsed across groups, automation use was highest with 28 ships (M = 83%, SD = 37%), slightly lower with 15 ships (M = 78%, SD = 40%), and lowest with two ships (M = 45%, SD = 49%). These differences were significant in a repeated-measures ANOVA ($F(2,169) = 19.14$, $p < .001$, $\eta_p^2=0.18$). Follow-up t-tests indicated that the difference in automation use between 2 and 15 ships (45% and 78%) was highly significant ($t(127)=5.38$, $p < .001$, $d = 1.01$), while that between 15 and 28 ships was not significant ($t(129) = -0.87$, $p = .38$, $d = .16$). As shown in Table 2, there was a common trend between the two variables: from 2 to 15 ships there was both a large increase in perceived difficulty and a large increase in automation use. But between 15 and 28 ships the increase in perceived difficulty, while significant, was much smaller, and while there was also an increase in automation use, it was not significant. This is also in line with raw accuracy, with highest accuracy at 2 ships and no difference between accuracies at 15 and 28 ships. This does not align with accuracy above chance, however, as that difference was between 15 and 28 ships. The accuracy above chance metric would be difficult for participants to calculate mentally and use in their decision to use automation to improve performance. Together, these findings provide support hypothesis 2a, that automation use differs when perceptions of difficulty differ, and suggest that perceptions of performance, rather than objective performance, may influence use. These findings are expanded upon in the discussion.

As an exploratory analysis, participants who rated the manual condition of 28 ships more difficult than the 15 ship conditions were separated into a group and their automation use was analyzed. This group of 13 participants did show a significant difference in their difficulty ratings ($t(12) = 9.81$, $p < .001$, $d = 2.72$) but their automation use was not significantly different ($t(12) = -1.24$, $p = .23$, $d = 0.40$). However, this lack of a difference may be due to low power with 13 participants, rather than a true failure to reject the null. Additionally, the automation use of participants who rated 15 ships as a 3 or lower on the difficulty scale was compared against the automation use of participants who rated 15 ships as a 5 on the difficulty scale. This was not significantly different ($t(50) = -1.39$, $p = .17$, $d = 0.39$) but again the effect size suggests that with larger power, there may be a difference. However, the current results suggest that perception may not always drive automation use – a direct contrast to hypothesis 2a. Automation use and difficulty perceptions were then correlated through a Spearman's Rank Correlation due to non-normal data. This resulted in a very small correlation (r = -.009), again contradicting hypothesis two. These results, combined with the overall automation use results, indicate that perceptions of difficulty are not consistently related to automation use, and some other influence is driving the decision more strongly.

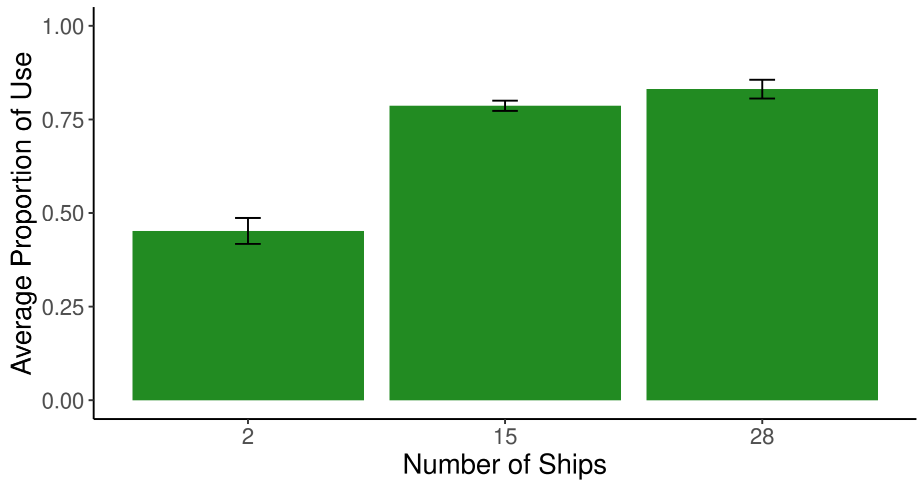

Figure 10. Automation use collapsed by ship number.

Trust and Self-Confidence

The Difference Between Trust and Self-Confidence

Previous experiments and literature suggest that trust and self-confidence in own performance influence the choice to use automation, such that higher trust and lower self-confidence lead to more automation use. The current experiment measured these metrics at each level of workload, unlike the previous experiment, allowing the mediation of workload in this relationship to be investigated. The difference between trust and self-confidence was a variable created by subtracting self-confidence from trust for each person. Overall, this difference was correlated with automation use (r = .32, p < .001), which aligns with both of the previous experiments here and supports the first hypothesis, that trust and self-confidence will be related to automation use.

Table 5. Correlations of automation use with trust, self-confidence and self-efficacy measures. One asterisk represents significance at the .05 level, and two asterisks represent significance at the .01 level.

	Trust	Trust-SC	Self Confidence	Self-Efficacy
Overall	0.12*	0.32**	-0.28**	-0.21**
15 ships	0.16*	0.31**	-0.18**	-0.09
2 ships	-0.21	0.05	-0.30*	-0.21
28 ships	0.31*	0.28*	-0.09	0.09

This correlation was similar in the blocks of trials with 15 ships (r = .31, p < .001) and 28 ships (r = .28, p = .05). However, although the correlation value in the 28 ship condition was similar, it was just significant, as was the slope of the line (slope estimate = .07, p = .05). The scatterplot indicated a slightly restricted range with most participants using the automation at high levels. Therefore, this correlation should be interpreted with caution, as it seems that the trust-self-confidence difference may not hold quite as well as very high workload levels. In the two ship condition, the correlation became much smaller and non-significant (r = .12, p = .45), suggesting that when the task was easy, automation use was determined by some other factor beyond the trust-self-confidence relationship. Self-confidence was similarly correlated, although negatively, with automation use (see Table 5). Trust was less strongly correlated with use, although still significantly (p = .04). These results suggest that self-confidence and the difference score of trust and self-confidence are the most strongly related to automation use, although trust does play a smaller role.

Table 6. Averages of trust, self-confidence, and the trust-self-confidence difference across the automation-choice blocks.

	2 Ships	15 Ships (2 anchor)	15 Ships (28 anchor)	28 Ships
Trust	3.20	3.21	3.34	3.49
Self-Confidence	3.31	2.43	2.31	2.17
Trust Minus Self-Confidence	-0.11	0.78	1.16	1.18

Table 6 presents the mean values of trust, self-confidence and the trust-self-confidence difference across the different conditions. The trust-self-confidence difference was essentially zero in the 2 ships condition (M = -0.11), since the trust and self-confidence averages were not significantly different from each other ($t(40) = 0.77$, $p = .44$, $d = 0.12$). In the 15 ship condition, the difference was positive (0.78) and trust was significantly higher than self-confidence ($t(171) = -9.76$, $p < .001$, $d = 1.02$). Between the 2 and 15 ship conditions, these average differences were significant ($t(40) = 5.58$, $p < .001$, $d = 0.86$). The trust-self-confidence difference was similarly high in the 28 ship condition (M=1.18) and not significantly different from the 15 ship condition (M=1.16; $t(44) = -0.10$, $p = .91$, $d = 0.01$). Note that self-confidence was only measured after automation-choice blocks, so comparing two and 15 ship, as well as 15 and 28 ship self-confidence levels requires between group comparisons. The self-confidence ratings at 15 ships between the two groups do not differ. These differences align with those of automation use – specifically, automation use differed between two and 15 ships, but not 15 and 28 ships, reflecting the identical pattern to the trust-self-confidence difference. The mirrored patterns suggest that automation use may be influenced by the trust-self-confidence difference.

Trust and Self Confidence in Relation to Performance

To better understand why trust and self-confidence might be related to automation use, self-confidence was correlated with unaided performance in the manual blocks. This would allow an understanding of how perceptions of performance (i.e. self-confidence) may differ from actual performance. Self-confidence was related to performance ($r = .34$ $p < .01$), although there is still a reasonable amount of unexplained variance, suggesting that while self-confidence is useful in predicting automation use, people are likely sensitive but not very well calibrated in their estimates of their own performance. This can also be seen in the self-confidence differences across workload conditions. While participants rated their self-confidence higher in the 2 ship condition ($M = 3.31$) than in the 15 ship condition (2.43; $t(40) = -4.83$, $p < .001$, $d = 0.90$), they did not rate their self-confidence differently in the 28 ship condition ($M=2.31$) from the 15 ship condition ($M=2.17$; $t(44) = -0.82$, $p = .41$, $d = 0.12$) even though accuracy was significantly lower. Note that self-confidence was only measured on automation-choice blocks and the comparison of 2-15 and 15-28 are from the two experimental groups, but the self-confidence ratings at 15 ships between the groups do not differ.

Trust was moderately related to performance ($r = .30$, $p =.004$). This links back to the idea that trust and dependence are not strongly related (Chapter 1), because if trust increased dependence, then this correlation would be stronger with better performance as a result of increasing use of the automation (due to the automation being more accurate than the human) as trust increased. Participants rated their trust no differently in the 2 ships condition ($M = 3.20$) than in the 15 ship condition (3.21; $t(40) = 0.22$, $p = 0.82$, $d = 0.02$). They did, however, rate their trust higher in the 28 ship condition (M=3.49) than in the 15 ship condition (M=3.34; $t(44) = -2.21$, $p = .03$, $d = 0.20$). Note that trust was only measured on automation-choice blocks and the comparison of 2-15 and 15-28 are from the two experimental groups, but the self-confidence ratings at 15 ships between the groups do not differ ($p=.10$). These differences are in opposition to the automation use changes across groups.

These findings point to the issue in perception-based measures. Self-confidence did not consistently change with performance changes, and trust changed across difficulty conditions even though the accuracy of the aid did not. These point to the inconsistencies that are seen when humans are asked to make judgements about themselves or their surroundings – a likely reason that designers prefer to stick with objective metrics.

New Measures of Trust and Self-Confidence

Due to reliability concerns in the subtraction of trust and self-confidence, two new questions were introduced to attempt to address this construct through more reliable measures. Participants rated their agreement with the statements “The automation is better than me at this task” and “I am better than the automation at this task”. These metrics showed very similar correlations with automation use as trust ($r=.17$) and self-confidence ($r=-.25$), respectively, but neither reached the level of correlation of subtracting self-confidence from trust.

Responses to the two questions were highly correlated ($r=.79$), and so the two were combined for a single response. Responses to “I am better than the automation at this task” were reverse coded, and these new responses were averaged with the responses to “The automation is better than me at this task” and correlated with overall automation use. This resulted in a small to moderate correlation of $r = .24$ ($p<.001$). This accounts for about 5% of the variance in automation use, which is half of the variance accounted for by the trust-self-confidence difference (r=.32). This suggests that the trust-self-confidence difference is somewhat better at predicting automation use, but with 90% of the variance still unaccounted for, there is still a large part of the equation missing, regardless of the metric used.

Self-Efficacy

Of interest was the way that self-efficacy may relate to automation use. Although it has been suggested that self-efficacy is different than self-confidence (Bandura, 1997), this has not been tested. Here, the two were closely correlated (r=.89), indicating that they may be interpreted by participants as the same metric or they are measuring the same construct. In relation to automation use, the were similarly (negatively) correlated. These results suggest that self-efficacy and self-confidence are very similar constructs in this context.

Regression Analysis

The correlational analyses and comparisons of averages suggest that perceptions of workload drive automation use, as does self-confidence and the trust-self-confidence. There are also some indications that task load might drive automation use – specifically the difference in use between 2 and 15 ships, but it is unclear why that difference does not also appear between 15 and 28 ships. To better parse out the individual effects of each of these metrics, perceived and objective workload, task difficulty perception, self-confidence, and trust were tested in a logistic regression.

First, perceived workload (ISA), perceived task difficulty, self-confidence, and trust were included in a non-interaction model on automation use across all ship numbers. Self-confidence was most strongly related to use (estimate(SE) = -0.53(0.14), $z = -3.64$, $p < .001$) where higher self-confidence predicts lower automation use. Trust was also related to use (estimate(SE) = 0.39(0.18), $z = 2.08$, $p = .03$), such that higher trust indicates more automation use, although this was a weaker effect than self-confidence. Subjective workload and difficulty were not significantly related to automation use. A second model was run to include the interaction of trust and self-confidence. This was also significant (estimate(SE) = -0.40(0.15), $z = -2.59$, $p < .01$). The interaction between trust, self-confidence and difficulty perceptions was not significant (estimate(SE) = -0.05 (0.12), $z = -0.41$, $p = .67$), suggesting that difficulty does not change the impact of the trust-self-confidence relationship on automation use. Lastly, a model was run with the trust-self-confidence difference variable and it was also a predictor of automation use (estimate(SE) = 0.89(0.39), $z = 2.27$, $p = .02$). Its interaction with difficulty was not significant (estimate(SE) = -0.12(0.11), $z = -1.12$, $p = .26$). Together, these results suggest that it is self-confidence and trust that influence automation use, not perceived workload or perceived difficulty.

Then, models were run to include task load. Task load here was operationalized by the number of ships on the screen – 2, 15, and 28. A model was run that included task load, perceived workload, perceived difficulty, self-confidence, and trust. Task load was a significant predictor of automation use (estimate(SE) = 0.06(0.02), $p = .004$), such that higher load does lead to more automation use. Self-confidence was also a predictor of use (estimate(SE) = -0.45(0.15), $p = .002$), such that higher self-confidence led to less automation use. Perceived workload and difficulty, and trust, were not significant predictors. A second model was run to include the interaction term of task load and self-confidence, but this was not significant ($p = .17$). To ensure that the effects of objective workload and self-confidence are unique, step-wise regression was conducted. Self-confidence alone influenced automation use (estimate(SE) = -0.46, $p < .001$), and when task load was added both effects were still significant (self-confidence: estimate(SE) = -0.36(.14), $p = .01$, objective workload: estimate(SE) = 0.06(0.02), $p < .01$). This, along with the lack of a significant interaction term, suggests that self-confidence and objective workload have unique influences on automation use.

In all, these regression models provide evidence that (a) self-confidence is consistently associated with individual choices in automation use, (b) perceived workload does not influence automation use, but task load measured as a by-product of performance does influence automation use, and (c) trust is not as strong of a predictor of automation use, although it may have some effects. The discrepancies between these results and those of the comparisons to accuracy and automation use are covered in the Discussion.

Discussion

This experiment aimed to explore if perceptions of difficulty, rather than objective measures of difficulty, influence automation use. To test this, high (28 ships) and low (2 ships) difficulty levels were used as anchors to change the perceptions of difficulty at 15 ships. Analyses indicated that perceptions of difficulty at 15 ships between the high and low difficulty anchors did not differ. Therefore, perceptions of difficulty were analyzed between groups of people. Participants who rated 28 ships as more difficult than 15 ships showed no significant difference in automation use between the two conditions, nor was there a significant difference in automation use between participants who rated 15 ships as more or less difficult, however these comparisons were exploratory and low-powered but had large effect sizes, which indicates that with better power there could be differences.

The perceptions of difficulty and ISA ratings across the three levels of task load did align with automation use, where large differences were seen between 2 and 15 ships (ISA: $d = 1.60$; use: $d= 1.01$) but not between 15 and 28 ships (ISA: $d= 0.23$; use: $d= .16$). This also aligns with raw accuracy scores, which participants would have known due to their score being shown on the screen across trials. However, this is in contrast to the accuracy above chance scores, which showed only a difference between 15 and 28 ships. This chance-corrected score, while more accurately representing performance, would be much more difficult for participants to mentally calculate as they completed the task and therefore is unlikely to have played a role in their understanding of their performance. These results suggest that when operators are choosing to use automation, part of their choice is being driven by their perception of workload, difficulty, and/or performance.

In contrast to the between-participant differences above, the logistic regression model suggested that perceptions of difficulty or workload did not predict automation use, but the task load (number of ships) did. However, this may be driven, in part, by the lack of a difference in perception of workload and difficulty between 15 and 28 ships compared to the change in raw accuracy. When looking to compare these results to other literature, the influence of task load here is in direct contrast to prior literature that suggested workload does not influence automation use (Riley, 1995; Lee & Moray, 1992; De Vries et al., 2003). It is, however, in line with the literature on dependence on static automation (Biros et al., 2004; Dixon et al., 2007; McBride et al., 2011; Xu et al., 2007). It may be that in the static automation literature, perceived and objective changes occurred in tandem, but that was not true in the few studies on discretionary automation.

Correlational analyses (Table 4) and the logistic regression model indicate the importance of self-confidence in the choice to use automation. Self-confidence was correlated at $r = -.28$ with automation use across all workload conditions. When self-confidence differed between conditions – specifically between two and 15 ships – automation use also differed. When self-confidence was similar – between 15 and 28 ships – automation use did not differ. This, combined with the significance of self-confidence in the regression model, indicates that self-confidence is a key part of deciding to use automation, even if self-confidence is not closely calibrated to actual performance, as seen here ($r = .20$). This aligns with the findings of the two prior experiments, but the logistic regression allows for the influence of workload and self-confidence to be explicitly separated, such that increases in automation use at higher workload are not only due to the decrease in self-confidence associated with a more objectively difficult task.

Prior literature has implicated the importance of self-confidence in automation use (Riley, 1995; Lee & Moray, 1992), but also includes the importance of trust in the automation. Here, trust in the automation was not a significant predictor of automation use in the logistic regression model when objective task load was included. It was a predictor when objective task load was not included, and this seems to indicate that the impact of trust on automation use was not independent from the change in trust across task load. Trust was slightly correlated with automation use overall. This aligns with the findings from the review on trust and dependence in Chapter 1 (trust-dependence correlation averaging $r = .11$), suggesting that trust alone does not have a strong relationship with automation use. Although the effects here are not strong, this is not to say that trust has no impact on automation use. It is thought that initial use of the automation is determined by trust, such that if a person has no trust at all in a system, they will not use it (Hoff & Bashir, 2015). Here, participants were forced to interact with the system in an initial block of trials, which provided a required amount of time to calibrate their trust before deciding if they wanted to use the aid or not. This may not mimic real world scenarios such as a self-driving car, where users who do not trust the system never turn it on – a clear case where trust is impacting use.

However, in scenarios where operators interact with a system or have some base level of trust, the current data suggest that trust should be considered in relation to self-confidence, possibly through a difference score. The trust-self-confidence difference correlated with automation use at $r = .32$, and was a significant predictor in the regression model. The trust-self-confidence difference also changed between 2 and 15 ships, but not between 2 and 28 ships, which aligns with the changes seen in automation use. These findings provide additional support for the previous findings in this dissertation and prior literature (Riley, 1995; Lee & Moray, 1992).

Although the correlation between the trust-self-confidence difference and automation use was numerically slightly larger than the correlation between only self-confidence and automation use, it does not account for much more variance. The argument has been made in measurement literature that difference scores, or metrics computed by subtracting two scores as done here in the trust-self-confidence difference, are less useful, reliable, and valid than the individual components that make up the difference score (Edwards, 1994). The current experiment suggests that the difference score, while useful, may not be better than only self-confidence. However, one experiment that has addressed the trust-self-confidence issue in static automation suggests it is trust, not self-confidence, that impacts automation reliance (Wiczorek & Meyer, 2019). Reliance and the choice to use automation are not the same, and so the difference in dependent variables and the static versus dynamic automation contexts are likely to play a role in the discrepancy between the Wiczorek and Meyer (2019) results and those found here. There is a need for future research to continue investigating this issue and look to understand if the difference score of trust and self-confidence or self-confidence alone uniquely explain the findings in static literature as well as dynamic literature.

Self-Efficacy and New Trust-Self-Confidence Measures

Tangential to the main goal of this experiment was an interest in the way self-confidence and self-efficacy were related. The two were highly correlated with each other, and showed similar correlations with automation use. This suggests that either participants interpret the questions in the same way, or the metrics are measuring the same construct. Typically, self-efficacy is thought to be more context specific than self-confidence, although in the current experiment the questions were very similar (see Appendix A for exact wording). Future research would benefit from testing general self-confidence measures against context-specific self-efficacy measures.

Also tested were two new complementary metrics that would allow trust and self-confidence to be measured through one unit, rather than through subtraction. These metrics accounted for 5% of the variance in automation use, which is nominally less than the 10% accounted for by the trust-self-confidence difference, suggesting that there is no basis to switch to a newer measure. However, both metrics still leave 90-95% of the variance in automation use unaccounted for, which indicates that neither can strongly predict automation use.

Conclusions

Perceptions of difficulty or task load seem to have an influence on automation use. Perceptions of difficulty aligned with automation use, and the objective difficulty metric of performance-above-chance accuracy did not align with use. These findings are not unequivocal, but do suggest that perceptions matter. Additionally, self-confidence and objective task load or difficulty impact use decisions. Specifically, when self-confidence increases, automation use decreases. This may occur in tandem with changes in workload or task difficulty, but it appears that self-confidence is a unique influence.

However, the trust-self-confidence difference and self-confidence alone only account for about 10% of the variance in the automation use each, suggesting there are other influencing variables. Trust also plays a role in automation use, but explains even less of the variance in total. However, the current study suggests that while automation use may increase if a task is more difficult, it is likely a result of personal perceptions of trust and self-confidence, not the perceptions of task difficulty. This is useful for designers to predict when operators will use automation across a variety of tasks, as understanding the operators' perceptions of their abilities versus perceptions of the automation can be used to ensure the automation is reliable enough to be used and is deemed necessary by the operator.

CHAPTER 6 - GENERAL DISCUSSION

Experimental Findings

This dissertation aimed to better understand how human operators decide to use automation. The literature on human-automation interaction is vast, but most of it focuses on static automation, or automated aids that are always present. As a result, the idea of "use" in the literature up to this point is more specifically a measure of dependence, or the extent to which an operator agrees with or depends upon the automation. The existing body of literature provides a number of variables that have consistently shown an influence on operator's dependence. Specifically, as tasks become more difficult, complacency (over-reliance on the automation) and general dependence tend to increase (Dixon et al., 2007; Parasuraman & Manzey, 2010). It has also been thought that trust in the automation plays a role in operator dependence, although a review of the literature suggests that this relationship is not very strong ($r = .13$; Chapter 1). Other variables such as automation reliability have been shown to impact dependence as well (Strickland et al., 2022).

As a result of decades of research, general trends of dependence on static automation can often be predicted – for example, increasing task difficulty will typically lead to more dependence. However, as automation becomes more ubiquitous in real world environments, its applications expand, and people have more choice surrounding their automation use. For this reason, understanding why people choose to use automation becomes crucial. Choosing the presence of automation, termed here "discretionary automation", lacks research on the way influences on choice may differ from those on dependence. There is some evidence that the variables in the static automation literature – particularly workload – may not hold in discretionary automation use (i.e., Riley, 1995), but these studies are outdated, lack strong variable control, and typically have low power. Through three new experiments, this dissertation provides evidence that workload influences discretionary automation use, as does trust, self-confidence, and the difference between the two metrics.

Task Difficulty and Workload

Previous literature suggests that objective workload, measured as task load or by performance, does not influence the choice to use automation (Riley, 1995; Harris et al., 1995), and subjective workload is not correlated with automation use (Chavaillaz et al., 2016). In contrast, there is evidence that static automation dependence is influenced by objective workload (Wickens & Dixon, 2007; Schwark et al., 2010), and so its effect was investigated here in an attempt to better understand the discrepancies and see if more variance in automation use could be explained.

In Experiment 1, both subjective workload of a concurrent task and task performance (primary and secondary) differed between task load conditions, but automation use did not change, providing evidence against the idea that subjective or objective workload on a concurrent task influences automation use. In contrast, in Experiment 2, task performance decreased and automation use increased with the highest task load version compared to the lowest task load of the primary task, which suggests that task load does influence automation use on a primary task. However, this experiment did not allow for a differentiation of subjective versus objective workload measures because the former was not assessed. In Experiment 3, there was evidence that automation use on the primary task aligned with both subjective workload and raw accuracy, but not the more objective metric of accuracy above chance. These differences in results across experiments may be due to differences in the types of workload. In Experiment 1, subjective workload was measured through a questionnaire about the trials as a whole, rather than only about the automated task. The change in difficulty for the secondary task influenced workload reports and performance, but participants likely did not recognize the impact of the increased difficulty on their primary automated task performance. In Experiments 2 and 3, the automated task was the only task that participants were asked to complete, which allowed for exploration of the influence of task load and subjective workload of the primary task on automation use.

Combined, the results of these three experiments suggest that when the workload of a concurrent, non-automated task is varied, it does not impact automation use even if it impacts performance or subjective workload ratings (Experiment 1). When the task load of the automated task was increased, there was a weak but consistent trend to increase automation use (Experiment 2: between 5 and 20 ships; Experiment 3: between 2 and 15 ships). It is unclear from the current results what drove the change in automation use across task load in Experiment 3. It may have been the change in self-confidence between 2 and 15 ships (Table 6), although the logistic regression suggests that was an independent factor. It may also have been the raw accuracy that participants saw, of which the differences mimicked automation use patterns. The logistic regression and correlational analyses revealed slightly weaker links between the experience of workload and automation use when examining effects across individuals. Looking across workload conditions, however, these results indicate that there is some impact of workload on automation use choices.

However, individual differences in objective performance of the operators did not seem to be related to automation use (Experiment 2). It would be optimal for operators with low accuracy to use the automation more often when the automation is more accurate than the unaided human. However, Experiment 2 indicated that neither unaided, aided, nor the extent to which automation improved performance, significantly correlated with automation use. Although indicating non-optimal decision making, these findings do align with the consistent finding from static automation literature that human operators tend to depend on imperfect automation at a level that reduces performance below that of the automation itself, indicating non-optimal dependence (Patton et al., *under review*; Bartlett & McCarley, 2017, 2020; Boeskemper et al, 2021).

It is worth noting that the objective measure of performance may not align with the participants' perceptions of their performance. Hoff and Bashir (2015) suggested that it is difficult for operators to use automation when they cannot tell how much it is helping them. Previous experiments indicated that the hostile ship task is difficult enough that participants may not take in additional information provided to them (Patton et al., 2022). Here, if participants were inaccurate in their estimates of performance, even though they were told when they were correct or incorrect in Experiments 2 and 3, or if they were unable to mentally compare their performance with and without the automation, it would be more difficult for them to optimize their use of the automation. This is likely related to self-confidence and becomes an issue in the real world, where operators may believe they are capable of completing a task without automation and it does not help them enough to warrant use, even if those beliefs are incorrect.

In conclusion, the current set of studies indicate that perceptions of difficulty or workload influence automation use, although on an individual level performance is not as predictive. This novel evidence for the importance of workload in discretionary automation indicates that the choice to use discretionary automation may not that different from the choice to depend on static automation.

Trust and Self-Confidence

Across all three experiments, the influence of self-confidence and trust on automation use was apparent. Individually, both trust and self-confidence are related to automation use, albeit in opposite directions. Specifically for trust, the correlation with automation use ($r = .12$ in Experiment 3 and $r = .16$ in Experiment 2) mimics that of the overall correlation between trust and dependence in the static literature ($r = .13$, Chapter 1). Although trust on its own does have an impact on automation use, it does not account for much of the variance. This provides an answer to the first research question – the extent to which trust impacts automation use, independent of workload or self-confidence, is limited.

Self-confidence tells a similar story. All three experiments showed a relationship between automation use and self-confidence, and although the correlation was slightly larger ($r = -.28$ in Experiment 3) than the trust correlation, the amount of variance explained is still quite small. Prior work in discretionary automation literature suggested it was the difference of trust and self-confidence (Lee & Moray, 1992) that influenced automation use the most. This was generally supported here, with the metric of self-confidence subtracted from trust showing moderate correlations in Experiments 2 and 3, and large differences across automation use groups in Experiment 1. In Experiment 3, the correlation of self-confidence and use was similar to that of the trust-self-confidence difference correlation. Trust and self-confidence, and the difference between the two, were also shown to be significant predictors of automation use in a logistic regression model, providing additional support for their importance. Therefore, it is unclear exactly to what extent the trust-self-confidence difference impacts automation use, as compared to only trust or self-confidence (Research Question 2).

The results of the current studies provide evidence for the importance of considering operator perceptions of their own abilities and those of the automation. The consistency of the influence of self-confidence across these three experiments indicates the importance of it – even though it has seldom been investigated in the human-automation interaction literature up to this point. When it has been studied in static automation literature, for example in one recent experiment, no impact was found of self-confidence automation reliance and compliance (Wiczorek & Meyer, 2019). The contradiction of this finding to those of the current experiment may be due to the difference in the choice to use automation as compared to relying or complying with automation.

Importantly, however, Wiczorek and Meyer (2019) provides an example of the way that self-confidence is seen as unimportant in the static automation literature. Rather, the focus of human-automation interaction literature, particularly in the last decade, has been on trust. This is not to say that trust is irrelevant - the current experiments here showed trust and use to be slightly correlated – but its average effect is small (d = 0.33, Chapter 1). Instead, these findings suggest that the current focus on trust in the literature may not be warranted if the goal is to understand or predict automation use or dependence. It seems that self-confidence should be considered – although it may benefit from a concurrent consideration with trust.

Self-Confidence and the Role of The Operator

The consistent impact of self-confidence across all three experiments suggests that it is a variable worth including in future studies or applied settings when predicting automation use. However, its importance here and in the studies of Lee and Moray (1992) and Riley (1995) should be considered in context. In discretionary automation studies in particular, but also often with other types of automation, participants are often asked to do a single task and engage with automation. In these scenarios, the participant typically has the mental capacity available to determine if they want the automation, and further decide how much to rely on the automation. Specifically, the point of these tasks is to introduce the automation, and so the role of the participant does not change much. Yet, Sheridan (1995) suggested that when automation is introduced, the role of the human changes from active participant to passive supervisor or monitor. The human is expected to oversee the automation and step in when it fails, but is no longer actively involved in the process of completing the task. In single task experiments, such as Experiments 2 and 3 here, the participant is still involved in the task because they have no other role or expectation. When we consider automation use in situations such as self-driving cars, however, the role of the operator does change. Drivers may be more likely to turn on the automation not as a result of their perceived inability to drive safely, but because they would prefer to take on the role of monitoring so they can eat breakfast or apply makeup on the way to work.

The argument may be made that multi-tasking experiments or higher levels of automation (Sheridan & Verplank, 1978) better capture the change in role. For example, introducing automation that only requires a veto of an action rather than the information processing automation in the current experiments may better capture the role change that occurs in self-driving vehicles. Additionally, changing the priority of a to-be-automated task alongside other necessary tasks can mimic these real world scenarios, such that if a non-automated task has higher priority, it may encourage automation use. Future research should investigate these issues and at that point, self-confidence must be measured and correlated to understand how its effect may change or become extinct as the role of the operator changes.

In the current studies, self-confidence was one of the best predictors of automation use, and yet only explained about 10% of the variance in automation use. It is clear there are other pieces of the puzzle – it is likely that the role of the operator is one of them.

Workload Measurement

The variable findings of the impact of workload across the three studies presented here lend themselves to a brief discussion of the measurement of workload. A potential issue when measuring workload subjectively is in its retrospective nature. Using questionnaires like the NASA-TLX or ISA require participants to think back on their experience and then respond accordingly. When workload is assessed directly after a trial or experiment, it can be argued that the recollection of the experience is still new enough to not be misremembered. However, in high workload tasks, it may be difficult for a participant to record their mental capacity in a task accurately. Especially in a task like that in Experiments 2 and 3, where previous research has suggested that the task is difficult enough to prevent participants from using additional helpful information to improve performance (Patton et al., *under review*), it may be that at higher numbers of ships, participants cannot accurately acknowledge their workload. This may be part of the reason for a lack of a change in subjective workload measurement from 15 to 28 ships in Experiment 3.

Summary

The current set of studies suggest that the active choice to turn automation on or off is similar in some ways to the choice to depend on static automation. Static automation sees an increase in dependence as objective task difficulty increases (Dixon & Wickens, 2006), and that is true here in discretionary automation, as long as task difficulty changes on the primary automated task. In contrast, the current results also implicate the importance of the difference between trust and self-confidence, and the order in which automation is presented (manually completing the task first increases automation use) – an effect not driven by trust and self-confidence. The perceptions of the operator about their own abilities are quite important in the choice to use automation, which is different from the objective metrics influencing dependence on static automation such as workload and automation reliability (i.e., Strickland et al., 2022).

Lastly, although tangential to the main focus of these studies, this work provided the first explicit test of the relationship between self-confidence and self-efficacy. The two had a very strong correlation and similar relationships with automation use, indicating that they likely are measuring the same construct. While further research should be conducted across other domains – especially because of the focus on domain specificity in Bandura's (2006) guide to creating self-efficacy scales versus the typical domain-general approach of self-confidence – it seems that the two may be able to be used interchangeably.

Applications

The strong indication that trust and self-confidence together drive automation use is an important finding for real-world automation contexts. This indicates that proper use of discretionary automation systems may be less about ideal design and more about user perceptions of the automation and the scenario. With the rise of ChatGPT, for example, we may be able to predict when students will conduct academic misconduct with ChatGPT – if a student feels low confidence in their ability to complete an assignment of high quality, but believes ChatGPT to be trustworthy and capable, they are more likely to use it.

The lack of consistent influence of task load on automation use is also an important consideration for real world scenarios. Automation is often implemented in scenarios where humans may struggle to perform at a level acceptable for the specific context. If operators are not sensitive to the need for automation in a more difficult environment, allowing them to control when the automation is present may result in disuse or misuse, which would lead to less than optimal performance. In these contexts, it may be better to provide always-present automation to encourage calibrated use of the automation. Alternatively, as the results from Experiment 1 suggest, training operators in difficult scenarios without automation first may increase automation use.

As mentioned previously, the focus on perceptions from operators of their own capabilities and the automation's capabilities are particularly important. This provides evidence for the need for designers to care about constructs like trust. An operator who does not trust a system will not be likely to turn it on, and even less likely if they feel they can complete the task themselves. For designers who intend for their product to be used, this quickly creates an issue. Additionally, the way operators perceive a task or automated aid can impact their overall perceptions of the environment. If an operator does not trust the automation but also does not believe they can complete a necessary task, this may cause stress or discomfort in an operator as they decide how to approach the situation. Considerations of operator perceptions can improve performance by improving calibrated automation use, but can also encourage happier workers.

Limitations

This work used a simple laboratory paradigm with novice participants to answer applied questions, which always brings limitations. It would be useful to replicate these findings in more applied settings, such as automated driving simulators. Additionally, the level of automation across all of these studies was very similar, with the automation providing a suggestion for a diagnosis or decision but the operator still making final decisions. This may limit the generalizability of the findings, as it is not yet known if different levels or degrees of automation would change the way operators choose to use them. Lastly, the changes in type of workload from Experiment 1 (concurrent task load) to Experiments 2 and 3 (workload of the primary task) may be part of the reason for differences in findings, and future work would benefit from comparisons of these types of workload in the same task.

Future Work

The current experiments set the foundation for future work into understanding how to encourage more calibrated perception of difficulty, as well as other variables that influence automation use and the extent of their influence. Specifically, future work should look at automation reliability, which is known to influence both trust and static automation dependence, to understand its influence on discretionary automation use. Some research has begun to investigate this, such as De Vries et al. (2003) and Moray et al., (2000), but more recent replications and investigations into the interaction of accuracy with other variables becomes necessary. Additional variables such as degree of automation (Parasuraman et al., 2000) should also be investigated to better understand how the role of the operator influences automation use. Future research into these types of variables can be combined with the foundation created in the current research to create a model of automation use.

Future work should also begin to explore what occurs after the automation is turned on. It is currently unknown how operator behavior toward the automation compares in static automation settings to “automation-on” discretionary settings. Specifically, there would be an interest in understanding how calibrated reliance and compliance are with the automation’s reliability once it is turned on, as compared to when it is always present.

With the understanding of the importance of self-confidence, researchers and designers can begin to look to understand how to promote decisions surrounding automation use based on other factors as well. Self-confidence is often over-estimated (Stone, 1994) and thus while it may predict actual automation use, it may not push behavior toward optimal automation use. It becomes important then to understand how to calibrate self-confidence and how to encourage operators to make choices surrounding automation use based on other factors, such as improvement to performance with automation.

Toward a Better Model

In the static automation literature, there is a general consensus that automation reliability, workload, and trust influence automation use. A similar consensus exists in the adaptable and adaptive automation literature, although the body of literature is smaller. Yet, these ideas are missing key pieces of the story. The current experiments indicate the importance of self-confidence, a metric that is seldom discussed in the human-automation interaction literature. Human perceptions matter, yet they are often ignored, or treated as an outcome rather than a predictive variable. Rather, the decision to implement automation is often based on “who is better” – an idea that dates back to Fitts’ MABA-MABA list (1951), which lays out the tasks that machines and men are better at, respectively. In some ways, this is valid. Implementing automation that performs worse than an unaided human is unlikely to be helpful, and may even lower performance (Wickens & Dixon, 2007), but even if the automation can objectively perform better, considerations around the human operator are still necessary.

The idea of “human-centered automation” is central in Human Factors Psychology (Sheridan, 1995; Parasuraman & Wickens, 2008), yet the idea of “human-centered” seems to drifting toward a focus on changes and improvements in human behavior due to the implementation of automation, rather than on the ways that we can support humans in their current state. This is not to say that understanding the impact of workload, reliability, prior experience, and trust are unimportant, but rather, they do not provide a clear picture. In Muir and Moray’s (1996) seminal work on automation, they say that trust in automation matters not because of its effects on dependence, but because it impacts the way the operator experiences the task and context. Low trust and forced high dependence can create stress and uneasiness in the operator, leading to other issues in performance and job satisfaction. Yet, newer literature focuses on how to predict or measure trust (i.e., Tenhundfeld et al., 2020), rather than on the impacts it has on the human operator.

This work aims to bring some of the focus of the research back to the experience of the human, and the impact that has on performance and automation use. In Figure 1, I suggested that workload, accuracy, trust and prior experience all have a positive influence on the choice to use automation. This is still true, but self-confidence must now be included as a negative influence. Future work should aim to build upon this model, looking to understand how the role of the human impacts automation use and dependence. Other aspects of human cognition, such as understanding of the automation (i.e., transparency or mental model), and preference or salience of other tasks, need to continue to be investigated and included in this model.

Future work must also look to remedy the issues that arise when testing applied questions in a laboratory setting. As mentioned earlier, changing the role of the human in these simple laboratory paradigms is inherently different than the way a human operator becomes a supervisor in a self-driving car or other automated system. While the level of control that a laboratory affords is useful in early stages of experimentation, the models created must hold up in real-world settings where variance and noise are abundant. The end goal is to accurately predict both automation use, and overall impacts to performance not only in the lab, but in real world scenarios. Without expanding the focus of influencing variables beyond current considerations, this simply will not be possible.

Conclusions

This dissertation set out to answer four main research questions:

1. To what extent does trust impact the decision to use automation, independent of workload or self-confidence?

Trust appears to be related to automation use, although the results of Experiment 3 suggest it is not an independent contributor. However, it does correlate with dependence and use in some ways. Specifically, if trust is very low, use will be very low.

2. To what extent does the relationship between trust and self-confidence impact the decision to use automation, and is this a better predictor than trust alone?

The difference between trust and self-confidence was a consistent predictor of automation use across all three experiments, and appeared to be more closely related to automation use than trust alone. However, the difference scores have their own statistical problems, and it is less clear that the difference score is better than self-confidence alone. It is clear, however, that self-confidence should be considered when predicting automation use.

3. Does objective task difficulty influence the choice to use automation, or is the perception of task difficulty a stronger influence?

Objective task difficulty does appear to influence the choice to use automation, in some scenarios. As with most human-centered research, the answer is not clear cut, but there is consistent evidence in Experiments 2 and 3 that large changes in objective workload, measured by task load or primary task performance, on the automated task can influence the choice to use automation. Importantly, perceived workload is also an influence, with the findings of Experiment 3 indicating that automation use decisions align best with perceptions of performance or difficulty.

4. To what extent does the order in which an operator completes a task manually versus with automation impact a later decision to use automation?

Manual completion or training on a task leads to higher automation use – an unexpected finding. This was not driven by the trust-self-confidence difference, but future work should look to understand how it may interact with self-confidence alone. This has important implications for real-world applications in training for automated systems.

The current research takes a deeper look at how perception of a task impacts automation use. This dissertation aimed to better understand the relative influence of self-confidence, trust, task difficulty, and prior experience with the automation on the choice to use automation. The difference between self-confidence and trust is a consistently influential factor. As the workload, whether objective or subjective, of a primary automated task increases between conditions, automation use also increases. This work is theoretically important for creating a model of the probability of automation use, and brings up the importance of considering human perception in automation research.

REFERENCES

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

Barg-Walkow, L. H., & Rogers, W. A. (2016). The effect of incorrect reliability information on expectations, perceptions, and use of automation. Human factors, 58(2), 242-260.

Biros, D. P., Daly, M., & Gunsch, G. (2004). The influence of task load and automation trust on deception detection. Group Decision and Negotiation, 13(2), 173-189.

Bliss, J. P., Gilson, R. D., & Deaton, J. E. (1995). Human probability matching behaviour in response to alarms of varying reliability. Ergonomics, 38(11), 2300-2312.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, *84*(2), 191.Bandura, 2006;

Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on psychological science*, *1*(2), 164-180.

Bartlett, M. L., & McCarley, J. S. (2017). Benchmarking aided decision making in a signal detection task. *Human factors*, *59*(6), 881-900.

Bartlett, M. L., & McCarley, J. S. (2021). Ironic efficiency in automation-aided signal detection. *Ergonomics*, *64*(1), 103-112.

Boskemper, M. M., Bartlett, M. L., & McCarley, J. S. (2022). Measuring the efficiency of automation-aided performance in a simulated baggage screening task. *Human factors*, *64*(6), 945-961.

Bowden, V. K., Griffiths, N., Strickland, L., & Loft, S. (2021). Detecting a single automation failure: The impact of expected (but not experienced) automation reliability. *Human Factors*, 00187208211037188.

Burrell, A. M., Allan, J. L., Williams, D. M., & Johnston, M. (2018). What do self-efficacy items measure? Examining the discriminant content validity of self-efficacy items. *British Journal of Health Psychology*, *23*(3), 597-611.

Calhoun, G. (2022). Adaptable (Not Adaptive) Automation: Forefront of Human–Automation Teaming. Human Factors, 64(2), 269-277.

Carswell, C. M., Clarke, D., & Seales, W. B. (2005). Assessing mental workload during laparoscopic surgery. *Surgical innovation*, *12*(1), 80-90.

Chancey, E. T., Bliss, J. P., Proaps, A. B., & Madhavan, P. (2015). The role of trust as a mediator between system characteristics and response behaviors. Human factors, 57(6), 947-958.

Chancey, E. T., Bliss, J. P., Yamani, Y., & Handley, H. A. (2017). Trust and the compliance–reliance paradigm: The effects of risk, error bias, and reliability on trust and dependence. Human factors, 59(3), 333-345.

Chavaillaz, A., & Sauer, J. (2017). Operator adaptation to changes in system reliability under adaptable automation. Ergonomics, 60(9), 1261-1272.

Chavaillaz, A., Schwaninger, A., Michel, S., & Sauer, J. (2019). Expertise, automation and trust in X-ray screening of cabin baggage. Frontiers in psychology, 10, 256.

Chavaillaz, A., Wastell, D., & Sauer, J. (2016). System reliability, performance and trust in adaptable automation. Applied Ergonomics, 52, 333-342.

Chavaillaz, A., Schwaninger, A., Michel, S., & Sauer, J. (2019). Expertise, automation and trust in X-ray screening of cabin baggage. *Frontiers in psychology*, *10*, 256.

Chien, S. Y., Lewis, M., Sycara, K., Liu, J. S., & Kumru, A. (2018). The effect of culture on trust in automation: reliability and workload. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *8*(4), 1-31.

Chien, S. Y., Lewis, M., Sycara, K., Kumru, A., & Liu, J. S. (2019). Influence of culture, transparency, trust, and degree of automation on automation use. *IEEE Transactions on Human-Machine Systems*, *50*(3), 205-214.

Chiou, E. K., & Lee, J. D. (2023). Trusting automation: Designing for responsivity and resilience. Human factors, 65(1), 137-165.

Cronbach, L. J. (1970). *Essentials of psychological testing* (3rd ed.). New York: Harper & Row.

Davenport, R. B., & Bustamante, E. A. (2010). Effects of false-alarm vs. miss-prone automation and likelihood alarm technology on trust, reliance, and compliance in a miss-prone task. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 54(19), 1513-1517.

De Vries, P., Midden, C., & Bouwhuis, D. (2003). The effects of errors on system trust, self-confidence, and the allocation of control in route planning. International Journal of Human-Computer Studies, 58(6), 719-735.

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(3), 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*(4), 564-572.

Dorfman, D. D. (1969). Probability matching in signal detection. Psychonomic Science, 17(2), 103-103.

Du, N., Huang, K. Y., & Yang, X. J. (2020). Not all information is equal: effects of disclosing different types of likelihood information on trust, compliance and reliance, and task performance in human-automation teaming. *Human factors*, *62*(6), 987-1001.

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(6), 697-718.

Edwards, J. R. (1994). Regression analysis as an alternative to difference scores. *Journal of management*, *20*(3), 683-689.

Fitts PM (1951). Human engineering for an effective air navigation and traffic control system. National Research Council, Washington, DC.

Foroughi, C. K., Devlin, S., Pak, R., Brown, N. L., Sibley, C., & Coyne, J. T. (2021). Near-perfect automation: Investigating performance, trust, and visual attention allocation. *Human factors*, 00187208211032889.

Forster, Y., Hergeth, S., Naujoks, F., Beggiato, M., Krems, J. F., & Keinath, A. (2019). Learning to use automation: Behavioral changes in interaction with automated driving systems. Transportation research part F: traffic psychology and behaviour, 62, 599-614.

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Inexperience and experience with online stores: The importance of TAM and trust. IEEE Transactions on engineering management, 50(3), 307-321.

Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., ... & Economou, E. (1999). An anchoring theory of lightness perception. Psychological review, 106(4), 795.

Grier, R., Wickens, C., Kaber, D., Strayer, D., Boehm-Davis, D., Trafton, J. G., & St. John, M. (2008, September). The red-line of workload: Theory, research, and design. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 52, No. 18, pp. 1204-1208). Sage CA: Los Angeles, CA: Sage Publications.

Harris, W. C., Hancock, P. A., Arthur, E. J., & Caird, J. K. (1995). Performance, workload, and fatigue changes associated with automation. The International Journal of Aviation Psychology, 5(2), 169-185.

Herdener, N., Wickens, C. D., Clegg, B. A., & Smith, C. A. P. (2016). Overconfidence in projecting uncertain spatial trajectories. Human factors, 58(6), 899-914.

Hergeth, S., Lorenz, L., & Krems, J. F. (2017). Prior familiarization with takeover requests affects drivers' takeover performance and automation trust. Human factors, 59(3), 457-470.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. Human factors, 57(3), 407-434.

Horrey, W. J., Lesch, M. F., & Garabet, A. (2009). Dissociation between driving performance and drivers' subjective estimates of performance and workload in dual-task conditions. Journal of safety research, 40(1), 7-12.

Huegli, D., Merks, S., & Schwaninger, A. (2022). Benefits of decision support systems in relation to task difficulty in airport security X-ray screening. *International Journal of Human–Computer Interaction*, 1-16.

Hussein, A., Elsawah, S., & Abbass, H. A. (2020). Trust mediating reliability–reliance relationship in supervisory control of human–swarm interactions. Human Factors, 62(8), 1237-1248.

Hutchinson, J., Strickland, L., Farrell, S., & Loft, S. (2022). The perception of automation reliability and acceptance of automated advice. Human Factors, 00187208211062985.

Jian, J. Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. International journal of cognitive ergonomics, 4(1), 53-71.

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

Kaplan, A. D., Kessler, T. T., Brill, J. C., & Hancock, P. A. (2023). Trust in artificial intelligence: Meta-analytic findings. *Human factors*, *65*(2), 337-359.

Karpinsky, N. D., Chancey, E. T., & Yamani, Y. (2016, September). Modeling relationships among workload, trust, and visual scanning in an automated flight task. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 60(1), 1550-1554.

Klein, G. (2008). Naturalistic decision making. *Human factors*, *50*(3), 456-460.

Kohn, S. C., de Visser, E. J., Wiese, E., Lee, Y. C., & Shaw, T. H. (2021). Measurement of trust in automation: A narrative review and reference guide. Frontiers in psychology, 12, 604977.

Koustanaï, A., Cavallo, V., Delhomme, P., & Mas, A. (2012). Simulator training with a forward collision warning system: Effects on driver-system interactions and driver trust. Human factors, 54(5), 709-721.

Larrick, R. P., Burson, K. A., & Soll, J. B. (2007). Social comparison and confidence: When thinking you're better than average predicts overconfidence (and when it does not). Organizational Behavior and Human Decision Processes, 102(1), 76-94.

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

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human factors, 46(1), 50-80.

Longo, L., Wickens, C. D., Hancock, G., & Hancock, P. A. (2022). Human Mental Workload: A Survey and a Novel Inclusive Definition. *Frontiers in psychology*, *13*, 883321.

Lyons, J. B., & Stokes, C. K. (2012). Human–human reliance in the context of automation. *Human factors*, *54*(1), 112-121.

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: an integrative review. Theoretical Issues in Ergonomics Science, 8(4), 277-301.

McBride, S. E., Rogers, W. A., & Fisk, A. D. (2011). Understanding the effect of workload on automation use for younger and older adults. Human factors, 53(6), 672-686.

Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J. D., Johansson, E., ... & Keinath, A. (2019). The "Out-of-the-Loop" concept in automated driving: proposed definition, measures and implications. Cognition, Technology & Work, 21, 87-98.

Merkle, C., & Weber, M. (2011). True overconfidence: The inability of rational information processing to account for apparent overconfidence. Organizational Behavior and Human Decision Processes, 116(2), 262-271.

Merritt, S. M., Lee, D., Unnerstall, J. L., & Huber, K. (2015). Are well-calibrated users effective users? Associations between calibration of trust and performance on an automation-aided task. *Human Factors*, *57*(1), 34-47.

Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human factors*, *46*(2), 196-204.

Monroe, S., & Vangsness, L. (2022,). The effects of task difficulty and stress on trust in an automated navigation aid. *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 66(1), 1080-1084. Sage CA: Los Angeles, CA: SAGE Publications.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological review*, 115(2), 502.

Moray, N., Rodriguez, D., & Clegg, B. (2000). Levels of automation in process control. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 44, No. 1, pp. 93-96). Sage CA: Los Angeles, CA: SAGE Publications.

Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. Ergonomics, 39(3), 429-460.

Norman, D. A., & Bobrow, D. G. (1979). Descriptions: An intermediate stage in memory retrieval. *Cognitive Psychology*, *11*(1), 107-123.

Nourani, M., King, J., & Ragan, E. (2020). The role of domain expertise in user trust and the impact of first impressions with intelligent systems. In Proceedings of the AAAI Conference on Human Computation and Crowdsourcing (Vol. 8, pp. 112-121).

Pak, R., Fink, N., Price, M., Bass, B., & Sturre, L. (2012). Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. Ergonomics, 55(9), 1059-1072.

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. Human factors, 52(3), 381-410.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human factors, 39(2), 230-253.

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

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

Patton, C. E., Warden, A. C., Pharmer, R. L., (*in press*). I spy differences in system reliability. Human Factors and Ergonomics Society Annual Meeting.

Patton., C. E., Wickens, C. D., Clegg, B. A., Smith, C. A. P., Noble., K. N.,(*under review*). Supporting detection of hostile intentions: Automated assistance in a dynamic decision making context.

Patton, C. E., Wickens, C. D., Noble, K. N., Smith, C. A. P., Clegg, B. A., (2022). How history trails and set size influence the detection of hostile intentions. *Cognitive Research: Principles and Implications.* 7(1), 1-12.

Pharmer, Rebecca L., Christopher D. Wickens, Benjamin A. Clegg, and C. A. P. Smith. "Effect of Procedural Elements on Trust and Compliance with an Imperfect Decision Aid." In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 65, no. 1, pp. 633-637. Sage CA: Los Angeles, CA: SAGE Publications, 2021.

Pop, V. L., Shrewsbury, A., & Durso, F. T. (2015). Individual differences in the calibration of trust in automation. Human factors, 57(4), 545-556.

Rice, S., & Geels, K. (2010). Using system-wide trust theory to make predictions about dependence on four diagnostic aids. *The Journal of general psychology*, *137*(4), 362-375.

Rice, S. (2009). Examining single-and multiple-process theories of trust in automation. The Journal of general psychology, 136(3), 303-322.

Riley, V. (1995). Issues Associated with Operator Use of Automation. SAE transactions, 1393-1397.

Robinson, 2001 Task complexity, task difficulty, and task production: exploring interactions in a componential framework. Applied Linguistics 221: 27–55.

Ross, J. M., Szalma, J. L., Hancock, P. A., Barnett, J. S., & Taylor, G. (2008, September). The effect of automation reliability on user automation trust and reliance in a search-and-rescue scenario. In proceedings of the human factors and ergonomics society annual meeting (Vol. 52, No. 19, pp. 1340-1344). Sage CA: Los Angeles, CA: Sage Publications.

Sauer, J., & Chavaillaz, A. (2018). How operators make use of wide-choice adaptable automation: observations from a series of experimental studies. Theoretical issues in ergonomics science, 19(2), 135-155.

Sauer, J., Kao, C. S., & Wastell, D. (2012). A comparison of adaptive and adaptable automation under different levels of environmental stress. Ergonomics, 55(8), 840-853.

Schwark, J., Dolgov, I., Graves, W., & Hor, D. (2010,). The influence of perceived task difficulty and importance on automation use. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 54, No. 19, pp. 1503-1507). Sage CA: Los Angeles, CA: SAGE Publications.

Schaefer, K. E., Chen, J. Y., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. Human factors, 58(3), 377-400.

Sheridan, T. B. (1995, October). Human centered automation: oxymoron or common sense?. In *1995 IEEE International Conference on Systems, Man and Cybernetics. Intelligent Systems for the 21st Century* (Vol. 1, pp. 823-828). IEEE.

Sheridan, T. B. (1989). Trustworthiness of command and control systems. In Analysis, Design and Evaluation of Man–Machine Systems 1988 (pp. 427-431). Pergamon.

Sheridan, T. B. (2019). Individual differences in attributes of trust in automation: measurement and application to system design. Frontiers in Psychology, 10, 1117.

Sheridan, T. B., & Verplank, W. L. (1978). Human and computer control of undersea teleoperators. Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab.

Skitka, L. J., Mosier, K. L., Burdick, M., & Rosenblatt, B. (2000). Automation bias and errors: are crews better than individuals?. The International journal of aviation psychology, 10(1), 85-97.

Spain, R. D., & Bliss, J. P. (2009, October). The effects of automation expertise and system confidence on trust behaviors. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, No. 4, pp. 344-348). Sage CA: Los Angeles, CA: SAGE Publications.

Stone, D. N. (1994). Overconfidence in initial self-efficacy judgments: Effects on decision processes and performance. *Organizational Behavior and Human Decision Processes*, *59*(3), 452-474.

Tattersall, A. J., & Foord, P. S. (1996). An experimental evaluation of instantaneous self-assessment as a measure of workload. Ergonomics, 39(5), 740-748.

Tenhundfeld, N. L., de Visser, E. J., Ries, A. J., Finomore, V. S., & Tossell, C. C. (2020). Trust and distrust of automated parking in a Tesla Model X. *Human factors*, *62*(2), 194-210.

Tikhomirov, L., Bartlett, M. L., Duncan-Reid, J., & McCarley, J. S. (2022). When the going gets tough: The efficiency of automation-aided signal detection declines with task difficulty.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. science, 185(4157), 1124-1131.

Vangsness, L., & Young, M. (2019). Central and peripheral cues to difficulty in a dynamic task. Human factors, 61(5), 749-762.

Voslinsky, A., & Azar, O. H. (2022). The effect of a reference point in task difficulty: How does a task that becomes irrelevant affect effort, feelings and perceptions. Judgment and Decision Making, 17(4), 797-815.

Wang, L., Jamieson, G. A., & Hollands, J. G. (2011). The effects of design features on users' trust in and reliance on a combat identification system. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 55(1), 375-379.

Wickens, C. D., Fitzgerald, N. J., Clegg, B. A., Smith, C. A. P., Orth, D., & Kincaid, K. (2020). Decision aiding for nautical collision avoidance: Trust, dependence, and implicit understanding of the decision algorithm. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 64(1), 1950-1954.

Wiczorek, R., & Manzey, D. (2010, September). Is operators' compliance with alarm systems a product of rational consideration?. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 54(19), 1722-1726.

Wiczorek, R., & Meyer, J. (2019). Effects of trust, self-confidence, and feedback on the use of decision automation. Frontiers in psychology, 10, 519.
Wiegmann, D. A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. Theoretical Issues in Ergonomics Science, 2(4), 352-367.
Xu, X., Wickens, C. D., & Rantanen, E. M. (2007). Effects of conflict alerting system reliability and task difficulty on pilots' conflict detection with cockpit display of traffic information. Ergonomics, 50(1), 112-130.
Yang, X. J., Schemanske, C., & Searle, C. (2021). Toward quantifying trust dynamics: How people adjust their trust after moment-to-moment interaction with automation. *Human Factors*, 00187208211034716.
Yeh, Y. Y., & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. Human factors, 30(1), 111-120.
Yuan Zhang, M., & Jessie Yang, X. (2017, September). Evaluating effects of workload on trust in automation, attention allocation and dual-task performance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 61, No. 1, pp. 1799-1803). Sage CA: Los Angeles, CA: SAGE Publications.
Ziv, G., & Lidor, R. (2021). Autonomy support and preference-performance dissociation in choice-reaction time tasks. Human Movement Science, 77, 102786.

APPENDIX A – QUESTIONNAIRES

Subjective Measure Questions

Trust Questions

From Lee and Moray (1992). All responded to on a scale 1-5, 1 being “not at all” and 5 being “completely”.

1. To what extent can the system’s behavior be predicted from moment to moment?
2. To what extent can you count on the system to do its job?
3. What degree of faith do you have that the system will be able to cope with all system states in the future?
4. Overall, how much do you trust the system?

Self-Confidence Question

Experiment 1:

1. How confident are you in your ability to complete this task without automation?

Experiments 2 and 3:

2. Overall, how confident are you in your ability to accurately identify the correct hostile ship without automation?

Experiment 3: ***Trust/Self-Confidence New Measure Questions***.

Participants ranked how much they agreed with these statements on a scale 1-5, 1 being “not at all” and 5 being “completely”.

1. “I am better than the automation at this task”
2. “The automation is better than me at this task”

Experiment 3: Self-Efficacy Question.

Participants ranked their certainty on a scale of 0 (not at all certain) to 100 (completely certain).

1. Rate how certain you are that you can correctly detect the hostile ship and behavior without the help of the automated aid.

Difficulty Question

Participants ranked the difficulty on a 1-5 scale, where 1 was "very easy" and 5 was "very difficult"

1. Overall, how difficult was this task?

ISA Workload Measurement

Participants were provided with the table below and asked "Based on the scale above, how would you rate this task?"

Level	Workload Heading	Spare Capacity	Description
5	Excessive	None	Behind on tasks; losing track of the full picture
4	High	Very little	Non-essential tasks suffering. Could not work at this level very long.
3	Comfortable Busy pace	Some	All tasks well in hand. Busy but stimulating pace. Could keep going continiously at this level.
2	Relaxed	Ample	More than enough time for all tasks. Active on ATC task less than 50% of the time.

1	Under-utilised	Very Much	Nothing to do. Rather boring.

APPENDIX B – WORKLOAD LITERATURE FOR TRUST-DEPENDENCE REVIEW

Chien et al., 2019	higher performance-based trust in the higher workload condition
Chien et al., 201	positive correlation in US participants between workload and trust, no dependence information
Chien et al., 2019	no difference in checking behavior across workload or reliability conditions
Chien et al., 2019	higher compliance in high workload and high reliability
Chien et al., 2019	highest reliance in low workload and low reliability conditions
Zhang & Yang, 2017	no effect of workload or reliability on trust
Zhang & Yang, 2017	less dwell time on the automated task as workload increased
Zhang & Yang, 2017	no effect of workload on performance
Sat et al., 2019	null effects of workload so all their other null effects are also not useful because they didn't successfully manipulate workload
Karpinksy et al., 2018	higher workload led to lower trust and less dwell time on the automated task - this was true in both experiments

APPENDIX C – RELIABILITY LITERATURE FOR TRUST-DEPENDENCE REVIEW

Bowden et al., 2021	higher trust with higher expected reliability, and miss of automation failure rate was higher in the higher expected reliability condition, which suggests higher reliance. Note: actual reliability was not varied, only expectations
Chancey et al., 2013	trust and RT were associated with reliability
Hussein et al., 2020	trust mediates the relationship between reliability and reliance
Sanchez et al., 2004	decreased reliability led to decreased trust and reliance, although they were not explicitly correlated in analyses
Du et al., 2020	higher trust, compliance and reliance with higher reliability
Chancey et al., 2017(6?)	higher trust with higher reliability, higher compliance for the false-alarm group with higher reliability, and higher reliance for the miss-group with higher reliance

APPENDIX D – MAIN FINDINGS FROM EACH EXPERIMENT

Experiment 1: Table of main findings. Listed as Mean (SD) in each cell.

	Difficult Condition	Easy Condition
Truck Dispatch Accuracy	57% (49%)	63% (48%)
Package Accuracy	81% (15%)	93% (12%)
Subjective Workload	3.55 (0.63)	3.33 (0.65)
Automation Use (% of people who turned it on)	60% (49%)	55% (48%)
Trust	5.35 (1.76)	5.49 (1.43)
Self-Confidence	6.55 (2.11)	6.33 (2.06)
Self-Confidence Minus Trust	1 (2.10)	1.03 (2.74)

	Manual First	Auto First
Truck Dispatch Accuracy	60% (11%)	59% (11%)
Package Accuracy	86% (4%)	87% (5%)
Subjective Workload	3.43 (0.53)	3.45 (0.54)
Automation Use (% of people who turned it on)	65% (38%)	51% (40%)
Trust	5.42 (1.51)	5.43 (1.56)
Self-Confidence	6.43 (1.88)	6.46 (1.94)
Self-Confidence Minus Trust	1.00 (2.74)	1.03 (2.10)

Experiment 2: Table of main findings. Listed as Mean (SD) in each cell.

	5 Ships	10 Ships	15 Ships	20 Ships	Overall
Accuracy	46% (49%	33% (47%)	35% (47%)	26% (S44%)	35% (47%)
Accuracy Above Chance	38% (25%)	28% (23%)	32% (23%)	24% (21%)	30% (23%)
Accuracy with Automation	54% (49%)	39% (49%)	43% (49%)	35% (47%)	43% (49%)
Accuracy without Automation	36% (48%)	28% (45%)	26% (43%)	16% (36%)	26% (44%)
Automation Use	52% (50%)	55% (49%)	59% (49%)	63% (48%)	67% (49%)
Trust	NA	NA	NA	NA	2.76 (0.79)
Self-Confidence	NA	NA	NA	NA	2.14 (0.94)

Correlations

	Self-Confidence	Trust-Self-Confidence Difference	Automation Use
Trust	.32**	.48***	0.25*
Self-Confidence		-0.67***	-0.08
Trust-Self-Confidence Difference			0.27*

Experiment 3: Table of main findings. Listed as Mean (SD) in each cell.

	2 Ships	15 Ships	28 Ships	2 - 15 Ships	28 - 15 Ships
Accuracy	74% (43%)	56% (49%)	49% (50%)	53% (50%)	58% (48%)
Accuracy Above Chance	50% (20%)	52% (19%)	48% (21%)	50% (19%)	55% (20%)
Perceived Difficulty	2.37 (1.08)	4.06 (0.91)	4.29 (.093)	3.97 (0.98)	4.15 (0.85)
Automation Use	45% (49%)	78% (40%)	83% (37%)	70% (43%)	83% (38%)

ISA	2.59 (0.96)	3.15 (0.88)	3.23 (0.84)	3.34 (0.86)	3.35 (0.91)
Trust	3.2 (0.90)	3.28 (0.77)	3.49 (0.73)	3.18 (0.85)	3.37 (0.68)
Self-Confidence	3.31 (1.01)	2.37 (0.99)	2.17 (1.22)	2.43 (1.03)	2.31 (0.95)
Trust Minus Self-Confidence	-0.11 (0.96)	0.91 (1.22)	1.18 (1.18)	0.78 (1.31)	1.16 (1.09)

Correlations

	Trust	Trust-SC	Self Confidence	Self-Efficacy
Overall	0.12^{*}	0.32^{**}	-0.28^{**}	-0.21^{**}
15 ships	0.16^{*}	0.31^{**}	-0.18^{**}	-0.09
2 ships	-0.21	0.05	-0.30^{*}	-0.21
28 ships	0.31^{*}	0.28^{*}	-0.09	0.09