Virtual Agent Interaction - Improving Cognitive Abilities and Trust for a Complex Visual Search Task

Heather H. Milecki
Wright State University
VIRTUAL AGENT INTERACTION – IMPROVING COGNITIVE ABILITIES AND TRUST FOR A COMPLEX VISUAL SEARCH TASK

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Industrial and Human Factors Engineering

By

HEATHER H. MILECKI
B.S., Embry-Riddle Aeronautical University, 2005

2015
Wright State University
Introduction: This thesis research examined a novel decision support aid ("Spatial Cue + Virtual Agent") on human performance in a simulated complex visual search task. Method: Participants in the “Control” condition did not receive support from an aid. Participants in the “Spatial Cue” condition received support from an aid in the form of a bounding box. Participants in the “Spatial Cue + Virtual Agent” condition received support from an aid in the form of a bounding box and a virtual agent. The aids’ reliability was held constant at one level, 70 percent. Image difficulty was based on clutter; clutter was manipulated by varying image white space. Results: The "Spatial Cue + Virtual Agent” improved participants’ Probability of Detection, sensitivity, trust, and confidence. Discussion: This study indicates that there is a potential to mitigate declines in automation trust by simply increasing aids’ humanness.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I.</strong> INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>RESEARCH STATEMENT</td>
<td>1</td>
</tr>
<tr>
<td>RESEARCH RATIONALE</td>
<td>2</td>
</tr>
<tr>
<td>RESEARCH IMPLICATIONS (THEORETICAL/BASIC AND PRACTICAL/APPLIED)</td>
<td>3</td>
</tr>
<tr>
<td>HYPOTHESES</td>
<td>3</td>
</tr>
<tr>
<td>DOCUMENT STRUCTURE</td>
<td>5</td>
</tr>
<tr>
<td><strong>II.</strong> BACKGROUND</td>
<td>6</td>
</tr>
<tr>
<td>HUMAN BEHAVIOR</td>
<td>6</td>
</tr>
<tr>
<td>HUMAN SOCIAL TRUST</td>
<td>7</td>
</tr>
<tr>
<td>VISUAL REPRESENTATION OF HUMAN EMBODIMENT</td>
<td>9</td>
</tr>
<tr>
<td>AUTOMATION</td>
<td>11</td>
</tr>
<tr>
<td>AUTOMATION RELIABILITY</td>
<td>14</td>
</tr>
<tr>
<td>AUTOMATION RELIANCE</td>
<td>15</td>
</tr>
<tr>
<td>AUTOMATION TRUST</td>
<td>16</td>
</tr>
<tr>
<td>VISUAL SEARCH</td>
<td>18</td>
</tr>
<tr>
<td>LITERATURE SYNTHESIS</td>
<td>21</td>
</tr>
<tr>
<td><strong>III.</strong> METHOD</td>
<td>23</td>
</tr>
<tr>
<td>PARTICIPANTS</td>
<td>23</td>
</tr>
<tr>
<td>MATERIALS</td>
<td>25</td>
</tr>
<tr>
<td>DESIGN</td>
<td>31</td>
</tr>
<tr>
<td>PROCEDURE</td>
<td>32</td>
</tr>
<tr>
<td><strong>IV.</strong> RESULTS</td>
<td>36</td>
</tr>
<tr>
<td><strong>V.</strong> DISCUSSION</td>
<td>45</td>
</tr>
<tr>
<td><strong>VI.</strong> CONCLUSION</td>
<td>53</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>55</td>
</tr>
</tbody>
</table>
APPENDIX A: INFORMED CONSENT .......................................................................................... 59
APPENDIX B: PRE-TEST SURVEY .......................................................................................... 61
APPENDIX C: POST-TEST SURVEY ....................................................................................... 64
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Participants Age</td>
<td>24</td>
</tr>
<tr>
<td>2. Highest Level of Education Completed</td>
<td>24</td>
</tr>
<tr>
<td>3. Alphabetic Character Image</td>
<td>26</td>
</tr>
<tr>
<td>4. Test Platform</td>
<td>28</td>
</tr>
<tr>
<td>5. Target Comprehension Check</td>
<td>33</td>
</tr>
<tr>
<td>6. Probability of Detection</td>
<td>38</td>
</tr>
<tr>
<td>7. Probability of False Alarm</td>
<td>39</td>
</tr>
<tr>
<td>8. Overall Sensitivity</td>
<td>40</td>
</tr>
<tr>
<td>9. Overall Response Bias</td>
<td>41</td>
</tr>
<tr>
<td>10. Overall Response Time</td>
<td>42</td>
</tr>
<tr>
<td>11. Confidence by Difficulty</td>
<td>44</td>
</tr>
<tr>
<td>12. Results Highlights</td>
<td>44</td>
</tr>
<tr>
<td>Table</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>1. Virtual Representations of Human Embodiment in Industry</td>
<td>10</td>
</tr>
<tr>
<td>2. Automation Levels</td>
<td>12</td>
</tr>
<tr>
<td>3. Signal Detection Theory Measures</td>
<td>29</td>
</tr>
<tr>
<td>4. Technology Attitude Survey Results</td>
<td>37</td>
</tr>
<tr>
<td>5. Trust Survey Results</td>
<td>42</td>
</tr>
<tr>
<td>6. Confidence Survey Results (Frequency (Percentage))</td>
<td>43</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENT

I wish to express my respect and sincere gratitude to my thesis advisor, Dr. Jennie Gallimore. Thank you for your expertise, encouragement, and most importantly your time over the years. To Dr. Subhashini Ganapathy and Dr. Alissa Golden, I am most appreciative for your agreeing to serve on my thesis committee. I am very much thankful to Wright State University’s Biomedical, Industrial, and Human Factors Engineering administrative staff, who provided essential academic planning oversight and direction.

I would like to extend sincere thankfulness to my family and friends who provided unwavering support and encouragement on this long journey. Finally, this thesis research is dedicated to my parents. They inspire me every day to live life with compassion and purpose.
I. INTRODUCTION

RESEARCH STATEMENT

The integration of higher-order human-like graphical elements and user interfaces is becoming more prevalent. Avatars, embodied conversational agents, humanoid avatars, talking head agents, virtual agents, smart virtual assistants, living actors, etc., are used in gaming, social networking, mobile applications, online shopping, and more. Computers, once regarded as novel devices, are now viewed to have traits in common with humans. Computers have been researched as cooperative partners, coaches, partners in cooperative dialogues, and secretaries (Nass, Fogg, & Moon, 1996). Human-like manifestations are viewed to create a more natural human-computer interaction by increasing users’ sense of interpersonal communication, intimacy, and immediacy with technology. Canadian philosopher and communication theorist Marshall McLuhan coined the expression, "The medium is the message." He implied that it is not so much what is being said but how, that is important; he believed that the vehicle through which a message is delivered may be more important than the message itself (Rupersburg, 2014).

Virtual agents serve as decision support aids; they help users understand equipment decisions, some even suggest potential solutions. In general, decision support aids are tools used to present only meaningful or pertinent information that originates from raw data (e.g., statistical algorithms, models) to users. Decision support aids can be presented visually, aurally, tactiley, or some combination thereof. Traditional visual decision support aids include indicator lights, spatial output (e.g., bounding boxes, coloration), text diagnostic (e.g., a “Target Present”)
message), coloration, and segmentation. Traditional auditory decision support aids include tones, complex sounds, and speech. Traditional tactile decision support aids include vibration, heat, or electrical stimulation. Virtual agents are a more novel decision support aid. This thesis investigated a virtual agent, presented as a human-like graphical head with low-guidance verbal prompts, for a complex visual search task in a domain of high complexity and high consequence.

RESEARCH RATIONALE

Image analysts who perform complex visual search tasks are assisted by traditional decision support aids. Examples include security screening operators searching for weapons in property, pipeline operators searching for pipeline leaks, air traffic controllers searching for relevant aircraft amongst irrelevant aircraft on a radar display, satellite imagery analysts searching for armored vehicle movement in hostile locations, radiologists searching for tumors in computed tomography scans, and manufacturing engineers searching for manufacturing defects. These image analysts are assisted by indicator lights, bounding boxes, tones, etc.

Such conventional decision support aids and the fact that automation is imperfect, may exacerbate negative perceptions of automation. These negative perceptions may be assuaged with virtual agents. According to Morrison (2009), “The purpose of creating a pseudo-social environment with avatars is to help humans forget that they are dealing with a machine which leads them to communicate with the avatar socially as though it were human” (p. 305). This thesis hypothesized that the presence of a virtual agent will improve participants’ complex visual search task effectiveness and efficiency, and aid acceptance (measured through the perception of trust and confidence), as compared to a traditional decision support aid.

Additionally, this research differed from the large array of recent work on virtual agents, which are commonly employed in domains of low complexity and low consequence. For example, online shopping virtual agents may provide purchasing recommendations or
instructions. In this case, shoppers are not required to strictly scrutinize the automated information and consequences are limited to lower sales or increased time for a successful transaction. Additionally, domains of high complexity and high consequence commonly use traditional decision support aids (e.g., indicator lights, spatial output, tones, and complex sounds); this research investigated a novel decision support aid in a simulated domain of high complexity and high consequence. This thesis represented a more psychological and information-processing approach for designing effective display modalities.

RESEARCH IMPLICATIONS (THEORETICAL/BASIC AND PRACTICAL/APPLIED)

This thesis reviewed the contributions of basic cognitive science (concepts, principles, extant literatures, paradigms, measures, statistical techniques, etc.); relevant domains include decision-making, visual search, automation, and Signal Detection Theory (SDT). While this research was not intended to progress knowledge in general cognitive science, it was intended to support the novel concept (virtual agent decision support aid presentation in domains of high complexity and high consequence) presented in this thesis. A virtual agent decision support aid engenders components of human face-to-face interaction. People self-identify with computers, apply gender stereotypes to computers, and conduct themselves with good manners [toward computers]; these responses are alike to human-human interaction (Nass, Fogg, & Moon, 1996). Other benefits of increased “humanness” design include improved morale and job satisfaction, and reduced fatigue and burnout. If the benefits of human face-to-face interaction can be extrapolated to trust in automation, communication interfaces may be more effective and efficient.

HYPOTHESES

This research aimed to validate the following main hypotheses:
1. The presence of a virtual agent will improve participants’ complex visual search task performance. To assess this hypothesis, the following performance measures were collected: Probability of Detection, Probability of False Alarm, sensitivity, and response bias. The null-hypothesis was that a significant difference in the above performance measures will not exist between the “Spatial Cue” and “Spatial Cue + Virtual Agent” conditions. The alternative hypothesis was that Probability of Detection will be greater, Probability of False Alarm will be lower, sensitivity will be greater, and response bias will be closer to zero for the “Spatial Cue + Virtual Agent”, as compared to the “Spatial Cue” condition.

2. The presence of a virtual agent will improve participants’ response time. To assess this hypothesis, participants’ response time (i.e., visual search duration) was collected. Response time was defined as: \( T_{\text{Start}} = \) Image is displayed and \( T_{\text{End}} = \) Decision is rendered or time elapsed. The null-hypothesis was that a significant difference in response time will not exist. The alternative hypothesis was that response time will be lower for the “Spatial Cue + Virtual Agent” condition, as compared to the “Spatial Cue” condition.

3. The presence of a virtual agent will be positively accepted by participants. To assess this hypothesis, participants responded to a four response category (1 indicated Strongly Disagree and 4 indicated Strongly Agree) Likert item on their confidence level after each trial (post-decision). Additionally, participants were surveyed on trust upon completing each condition. The null-hypothesis was that a proportional difference in confidence and trust will not exist between the “Spatial Cue” and “Spatial Cue + Virtual Agent” conditions. The alternative hypothesis was that confidence and trust will be greater for the “Spatial Cue + Virtual Agent”, as compared to the “Spatial Cue” condition.
DOCUMENT STRUCTURE

The structure of this thesis report is as follows. Section I presents the research statement and rationale; theoretical and practical research implications; hypotheses; and the document structure. Section II presents relevant scholarship pertaining to decision-making, visual search, automation, and other topics supporting the study’s rationale; definitions, universal concepts, and key findings are presented. Sections III, IV, V, and VI present the research methodology, results, results discussion, and conclusion respectively. Supplemental documents (e.g., surveys) are included in the appendices.
II. BACKGROUND

The purpose of this section is to provide a review of research relevant to the use of a virtual agent as a decision support aid for a complex visual search task. Research topics include: human behavior, human social trust, virtual representation of human embodiment, automation, automation reliability, automation reliance, automation trust, and visual search.

HUMAN BEHAVIOR

Human behavior is directed toward persons, places, objects, and events. It is influenced by many factors ranging from culture, to faith, to attitudes, to genetics. Erickson (2013) summarized daily examples of human behavior: conversing with family, friends, or colleagues; engaging in routine exchanges with familiar strangers at the bus stop; deciding not to stop at the store because the parking lot is jammed; and joining in a standing ovation even though the performance was not very enjoyable. Such behavior is displayed and received through verbal and non-verbal (e.g., eye contact) communication; this communication provides behavior control and planning, and ultimately influences actual behaviors.

This thesis research conceptually focused on human behavior as directed toward other humans and objects. Gratch, Okhmatovskaia, Lamothe, Marsella, Morales, Werf, and Morency (2006) stated that, “Humans respond to each other, engaging in non-conscious behavioral mimicry and back-channeling feedback. Such behaviors produce a subjective sense of rapport and are correlated with effective communication, greater liking and trust, and greater influence between participants” (p. 1). Humans also extend these behaviors to objects. They name their vehicles, talk nicely to electronics, and form emotional attachments to houses. Humans have a
willingness to anthropomorphize, to treat the inanimate as real; they are willing to ascribe such lifelike qualities in different and sometimes partial ways (Taylor, 2009).

Behavior toward a virtual agent encompasses behavior directed toward both real-world humans and objects. While virtual agents commonly represent real-world humans, they are only digital representations (computer objects). However, Erickson (2013) stated that human behavior can be shaped based on the belief that an audience exists. If real-world humans and virtual agents can be perceived similarly, behavior toward technology may be positively altered, possibly increasing user performance, engagement, and motivation with technology. Other researchers (Nass, Steuer, Tauber, & Reeder, 1993; Nass, Steuer, & Tauber, 1994; Reeves & Nass, 1996) agree that human-human and human-computer interactions are similar (Nass, Fogg, & Moon, 1996).

HUMAN SOCIAL TRUST

Numerous definitions exist for the term ‘trust’. Gambetta (1990) (as cited in Abdul-Rahman & Hailes, 2000) defined trust as the probability that an action, which affect others, is performed before it can be monitored. Boon and Holmes (1991) (as cited in Adams & Webb, 2002), defined trust as “A state involving confident predictions about another’s motives with respect to oneself in situations entailing risk” (p. 1). Misztal (1996) (as cited in Welch, Rivera, Conway, Yonkoski, Lupton, & Giancola, 2005), stated that “Trust consists of believing that the consequences of someone’s intended action will be appropriate from our own point of view” (p. 7). Costa, Roe, and Taillieu (2001) (as cited in Adams & Webb, 2002), defined trust as a “psychological state...based on expectations of others and on perceived motives and intentions in situations entailing risk with others” (p. 2). These definitions share two common themes: 1) Trust entails a level of risk and 2) Trust entails believing another’s actions are genuine, with respect to oneself.
This thesis hypothesized that a virtual agent [in a decision support aid capacity] will improve participants’ acceptance (measured by visual search performance and the perception of trust) of automation in a twofold manner. First that virtual agents represent humans; second that some benefits of human-human trust can be extrapolated to human-automation trust. Abdul-Rahman and Hailes (2000) stated that “Trust is a social phenomenon and any artificial model of trust must be based on how trust works between people in society” (p. 2). As such it is important to explore trust as it is most commonly known, between people.

Misztal (1996) (as cited in Welch, Rivera, Conway, Yonkoski, Lupton, & Giancola, 2005), stated that “Trust provides a crucial basis for social order by setting the most basic limiting conditions necessary for human interactions to continue” (p. 7). Trust between humans can be of low complexity and low consequence: A mother trusting her teenage son to purchase everything on her grocery list. It can also be of high complexity and high consequence: Special Operations Forces members trusting each other to fight to the death for others and mission success. Without trust, people would have to dedicate high mental resources to monitoring the behavior of others (Adams & Webb, 2002).

Common trust themes state that it is hard to earn and easy to lose, and that it may take years to establish and seconds to undo. Early stages of trust establishment involve predictions of another based on behavioral evidence; later stages are based on a coherent system of knowledge that has been formed over time. By the later stages, a personal history has developed and another’s behavior have been deemed trustworthy (Adams & Webb, 2002). This of course goes both ways, just as ‘low trusters’ do not trust another until there is clear evidence that trust is justified, ‘high trusters’ trust another until there is clear evidence that trust is not justified (Rotter, 1980; as cited in Welch, Rivera, Conway, Yonkoski, Lupton, & Giancola, 2005).
While this thesis focused on automation trust, individuals’ trust in people and organizations influence how individuals interact with technology (or automation). Siegrist (1999, 2000) (as cited in Siegrist & Cvetkovich, 2000) found that participants positively accepted gene technology when gene technology scientists and gene technology companies were [socially] trusted.

VISUAL REPRESENTATION OF HUMAN EMBODIMENT

The visual representation of human embodiment takes on many forms, the most well-known is an avatar. Avatar is a Hindu term that is commonly translated to "appearance" or "manifestation" in English. In a computing context, avatars are virtual humans controlled by a live participant, however the term has been incorrectly used to encompass many virtual representations of human embodiment. Other representations include embodied conversational agents, humanoid avatars, talking head agents, virtual agents, smart virtual assistants, living actors, etc.

Visual representations of human embodiment, commonly found in academia research, now flourish in the government and industry. In 2008, the Air Education & Training Command stated that the effectiveness of education, training, and experience is dependent on timing, location, and format. The U.S. Army and Air Force started to focus requirements on “customized learning, mass collaboration, push and pull learning systems, distributed learning opportunities, simulated and virtual technology, and visualization technologies” (p. 5) (Sottilare, 2009). Today, warfighters are exposed to artificial intelligence through computer-based tutors, virtual characters in games and other simulations, and expert decision support tools in their training environments today (Sottilare, 2009).

Additionally, visual representations of human embodiment in industry has boomed in recent years (refer to Table 1).
<table>
<thead>
<tr>
<th>Virtual Representation Role</th>
<th>Finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interviewer</td>
<td>Richman, Kiesler, Weisband, and Drasgow (1999) (as cited in Powers, Kiesler, &amp; Torrey, 2007) found that people reveal more in a computer interview than in a face-to-face interview. Sproull, Subramani, Kiesler, Walker, and Waters (1996) (as cited in Bauer &amp; Neumann, 2005) showed that participants attributed a higher degree of trustworthiness to a user interface that had a human face than to a text-based interface.</td>
</tr>
<tr>
<td>Career Counselor</td>
<td>Bauer and Neumann (2005) found avatars to take the position of a trust intermediary in electronic commerce; the avatars were found to establish and influence consumer trust toward the supplier.</td>
</tr>
<tr>
<td>Salesperson</td>
<td>Lester, Converse, Kahler, Barlow, Stone, and Bhogal (1997) (as cited in Piwek, 2007) showed that pedagogical agents can improve students’ perception of the learning experience.</td>
</tr>
<tr>
<td>Teacher</td>
<td>Gerhard, Moore, and Hobbs (2005) (as cited in Youngblut, 2007) found that virtual gallery museum visitors reported higher levels of presence when accompanied by an embodied conversational agent, as compared to visitors who experienced the virtual museum alone.</td>
</tr>
<tr>
<td>Docent</td>
<td>Ijsselsteijn, de Kort, Bonants, Westerink, and de Jager (2004) (as cited in Youngblut, 2007) found participants who experienced a virtual coach while cycling on a stationary exercise bicycle to give significantly higher scores on spatial presence and significantly lower scores on negative effects. Bickmore and Picard (2005) (as cited in Foster, 2007) studied long-term social-emotional relationships and embodied agents by using an agent that acted as an exercise advisor.</td>
</tr>
<tr>
<td>Coach</td>
<td>Choi, Miracle, and Biocca (2001) (as cited in Youngblut, 2007) found participants who experienced an advertising agent to have significantly higher spatial and social presence scores than other participants. The agent was also associated with significantly more favorable attitudes and behavioral intentions. A positive relationship between presence and the measures of advertising effectiveness were also found.</td>
</tr>
<tr>
<td>Advertising Agent</td>
<td>Cassell, Bickmore, Campbell, Vilhjálmsson, and Yan (2000) (as cited in Foster, 2007) used an embodied agent as a realtor that answered users’ questions</td>
</tr>
</tbody>
</table>
Finally, the use of visual representations of human embodiment is advocated by end-users. Witmer and Singer (1998) (as cited in Powers, Kiesler, & Torrey, 2007) reported that research participants frequently claimed that they did better on a task because of the strong sense of presence they experienced. Foster (2007) found an embodied agent to improve users’ satisfaction, engagement, and opinions toward a computer system. Durlach and Slater (1998) (as cited Fabri, Moore, & Hobbs, 2002) found avatars to increase the sense of community by becoming a genuine representation of underlying individuals, not only visually, but also within a social context.

**AUTOMATION**

There are numerous definitions for the term *automation* in the technical literature. Kelly, Boardman, Goillau, and Jeannot (2000) (as cited in Adams, Bruyn, Houde, & Angelopoulos, 2003), defined automation as a “device or system that accomplishes (partially or fully) a function that was previously carried out (partially or fully) by a human operator” (p. 11). Automation is initially based on a measurement (an observation) which is then compared to stored information about expected and unexpected characteristics. The compared data is then evaluated against a criterion value, or threshold. If the data is greater than or equal to a preset threshold, a signal is relayed to a human operator (Sorkin & Woods, 1985). As previously
discussed, these signals take on many forms, including virtual agents. For example, a virtual coach ‘coaches’ a trainee based on physical condition measurements that are collected via sensors placed on the [trainee’s] body and analyzed against stored information (e.g., standard heart rate measurements). If these analyzed measurements are below standard thresholds for gender, age, and weight, a virtual coach may instruct a trainee to increase his speed, resistance, etc., for a given physical activity.

Sheridan (1996) (as cited in Jamieson, Wang, & Neyedli, 2008), stated that “It is a common misconception that automation is introduced to replace human operators with the purpose of alleviating human errors” (p. 15). On the contrary, this relationship reflects more of a paired partnership where the success of automation depends on how well it is used by humans. An optimal human-automation partnership balances the strengths and limitations of humans and automation. Human strengths include historical or contextual knowledge (Sorkin & Woods, 1985), pattern recognition, deliberate decision making, and the ability to quickly adapt to challenges (Drury & Sinclair, 1983; as cited in Jiang, Khasawneh, Kaewkuekool, Bowling, Melloy, & Gramopadhye, 2003). Automation strengths include the capability to perform complex data-processing (Sorkin & Woods, 1985) and memory storage and retrieval (Kantowitz & Sorkin, 1987; as cited in Jiang, Khasawneh, Kaewkuekool, Bowling, Melloy, & Gramopadhye, 2003). Revisiting the previous example, while a virtual coach quickly analyzes physical condition measurements, it lacks historical knowledge such as medical history and family background.

Parasuraman, Sheridan, and Wickens (2000) (as cited in Cummings, 2004) present (refer to Table 2) levels of this human-automation partnership.

<table>
<thead>
<tr>
<th>Table 2 Automation Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automation Level</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
</tbody>
</table>

Moray, Inagaki, and Itoh (2000) (as cited in Adams, Bruyn, Houde, & Angelopoulos, 2003) further grouped these levels; primary control resides with the human in the first four levels, levels 5 to 7 reflect a more give-and-take relationship between the human and the machine, and primary control resides with the machine in levels 7 to 10.

According to Mosier and Skitka (1996) and Sheridan (2002) (as cited in Jamieson, Wang, & Neyedli, 2008), for most automated systems, humans are still required to monitor and supervise automation. This dynamic collaboration lends to more complex issues (e.g., automation reliance and automation trust) which may increase task complexity. Now humans have to determine the correctness of automation, in addition to how influencing factors impact its correctness (Cummings, 2004). To mitigate this downside of automation, engineers must consider operators’ mental workload, manual skill, situational awareness, etc., when designing and implementing automation (Bainbridge, 1983; Parasuraman, Sheridan, & Wickens, 2000; Skitka, Mosier, & Burdick, 2000; as cited in Jamieson, Wang, & Neyedli, 2008).
AUTOMATION RELIABILITY

While it is unrealistic to separate the human component from a human-automation partnership, this separation is important when speaking strictly of automation reliability. Adams, Bruyn, Houde, and Angelopoulos (2003) stated that “Automation reliability refers to the extent to which automation does the job that it was designed to do” (p. 58). It is an equipment-level measure in its true meaning and does not include system-level factors (end-users, job aids, procedures, the environment, and other elements designed to interact for some common purpose). Automation reliability is often defined by a value between 0 and 1, which is one minus the probability that the automation will fail (a potential target is not the actual target (a.k.a., false alarm) or a potential target is not identified (a.k.a., miss)). It is calculated per the following formula: Automation Reliability = (1 – (Probability of a False Alarm + Probability of a Miss)). Values closer to 1 indicate greater automation reliability.

As with any type of automation, failures are expected. Imperfect automation is due to a tradeoff between speed and accuracy. For example, image processing algorithms could detect all weapons in air travelers’ property, however increased hardware and software demands would be required, as well as processing durations. Economically, this is not feasible, as increased hardware and software demands require additional funding. Operationally, this is not feasible, as air travelers’ property would not be processed at an acceptable rate and travelers would miss their flights. It is important to note that imperfect automation does not imply useless automation. Several studies have shown that even when imperfect, [human] performance with [imperfect] automation remains higher than purely manual performance (Wickens & XU, 2002).

benefits at a reliability level less than 70 percent. Imperfect automation changes the way humans interact with automation, specifically users’ perception of the reliability of the automation. Various factors affect perceived automation reliability, including operators’ opinions on the expected reliability of the automation, whether automation failures are attributed to humans or automation, the type of error (false alarm or misses), and the difficulty of the decision [made by the automation] (Goh & Wiegmann, 2006). Evidently, increasing automation reliability is only half of the solution for improving human-automation system performance; the other half requires humans to perform their part of the task (Sanchez, 2006).

AUTOMATION RELIANCE

Humans ‘part of the task’ includes automation reliance. In basic terms, automation reliance refers to whether a human uses or does not use automation. This decision is not necessarily binary, as users may elect to use [or not use] automation periodically. Additionally, this decision is affected by physical (e.g., fatigue), mental (e.g., self-confidence), emotional (e.g., trust in automation), social (e.g., perceived risk), and operations (e.g., time constraints) factors (Dzindolet, Pierce, Beck, & Dawe, 1999; Lee & Moray, 1992; Mosier & Skitka, 1996; Riley, 1994; Riley, 1996; as cited in Jamieson, Wang, & Neyedli, 2008). The complexity of the decision to use or not use automation gives way to its problematic nature: humans fail to rely upon it appropriately.

The two extremes of automation reliance can be characterized as over-reliance and under-reliance, or misuse and disuse. Misuse occurs when individuals over-rely on automation (Lee & See, 2004; as cited in Jamieson, Wang, & Neyedli, 2008). In 1992, Airbus A320 pilots (Flight 148) incorrectly set the flight management system and failed to take manual control of the aircraft even as it crashed into terrain; 87 out of the 93 individuals on-board perished (Sparaco, 1995; as cited in Lee & See, 2004). As previously discussed, automation failures are
expected and humans must intervene at times in which automation is believed to be incorrect [for dynamic human-automation partnerships]. Another real-world example, in 1995, the Royal Majesty cruise ship grounded when crew failed to intervene for 24 hours when the automated navigation system malfunctioned (Lee & Sanquist, 2000; National Transportation Safety Board, 1997; as cited in Lee & See, 2004). These examples illustrate the automation reliability-reliance paradigm described by Sanchez (2006): Automation that is highly reliable, likely on aircraft and ships, lends itself to a high probability of correct outcomes; unfortunately it also indirectly leads itself to a low probability of a correct outcome when automation fails.

The other extreme of automation reliance, disuse, occurs when humans reject automation in preference for manual control (Lee & See, 2004; as cited in Jamieson, Wang, & Neyedli, 2008). During the Gulf War, soldiers were equipped with automation (combat identification aids) designed to assist them with identifying friendly, neutral, or adversarial individuals. Fearing the penalties (e.g., killing a fellow soldier or unarmed neutral individual), some soldiers turned off their combat identification aids, preferring self-reliance to properly identify individuals (Dzindolet et al., 2000; as cited in Jamieson, Wang, & Neyedli, 2008). Automation disuse requires humans to allocate more cognitive and physical resources to verify information, which, of course, contradicts the intended objective of automation.

AUTOMATION TRUST

Automation misuse and disuse can be reduced by attaining an appropriate level of automation trust. As with human social trust, numerous definitions exist for the term automation trust. Lee and See (2004) defined automation trust as humans’ willingness to believe that automation will help them achieve their goals in high-risk situations; humans’ willingness to believe information from automation or make use of its capabilities (Pasuraman & Miller, 2004); and humans’ willingness to rely on automation despite its risks (Muir & Moray,
The development of automation trust is similar to the development of human social trust. Automation must perform consistently in a manner familiar to human operators, and support humans’ ability to predict its future behavior (Sheridan, 1988; as cited in Adams, Bruyn, Houde, & Angelopoulos, 2003). These characteristics contribute to establishing and maintaining an appropriate level of automation trust.

Trust in automation is “rarely wholly internally consistent”; it is possible to trust automation in one condition and not another (Adams, Bruyn, Houde, & Angelopoulos, 2003). The preferred condition, when automation performs as intended and as expected, establishes trust. Muir (1994) (as cited in Adams & Webb, 2002) stated that humans will confer trust on an automated system when trust is consistently validated. More commonly, automation performs with acute failures. While trust will decline in this instance, it will also recover when automation performs reliably for an extended period of time (Lee & Moray, 1992; 1994; as cited in Sanchez, 2006). Finally, chronic automation failures may result in complete automation distrust. Humans will either return to performing the task manually or learn to accommodate the failures. As previously discussed, human strengths (e.g., ability to recognize patterns, make rational decisions) allow operators to understand automation failures (Itoh, Abe, & Tanaka, 1999; Lee & Moray, 1992; as cited in Lee & See, 2004).

Automation trust is at the forefront of improving automation reliance until higher levels of automation reliability can be achieved. Trust in automation is traditionally increased by improving the precision of automation output (e.g., bounding box tightness), selectively applying automation (i.e., reduce unnecessary automation), or informing users of automation capabilities. Merritt and Ilgen (2008) investigated participants’ performance on a complex visual search task across two conditions (high functioning machine and low functioning machine). Participants in the high functioning machine condition were primed or provided with
information to convey increased machine competence, responsibility, predictability, and dependability. These participants demonstrated increased trust, devoid of actual machine characteristics, as compared to participants in the low functioning machine condition.

This thesis investigated a more novel approach, simply increasing the humanness of automation (decision support aids) to increase automation trust. More specifically this research utilized a virtual agent to increase the humanness of automation. Research by Lerch, Prietula, and Kulik (1997) (as cited in Adams, Bruyn, Houde, & Angelopoulos, 2003) showed that participants trusted advice more when they believed that the advice was from a human, as compared to a computer. There is also a quickly accumulating body of evidence that, even though humans do not appear fooled by automation, they respond more positively to automation that exhibits human traits (Friedman, 1995; Friedman & Nissenbaum, 1997; Breazeal, 1999; Friedman, Khan, Howe, 2000; Prendinger & Ishizuka, 2001; DiSalvo, Gemperle, Forlizzi, & Kiesler, 2002; as cited in Nickerson & Reilly, 2004).

VISUAL SEARCH

The final literature topic, visual search, sets the context for this thesis research. Visual search is the process of searching an area for a known target (i.e., distinguishing a target from distractors). It occurs in everyday life, in both personal and professional contexts, and can be simple or complex. A personal example is searching for car keys on a disorganized work desk; the target car keys must be discerned from distractors such as binder clips, reading glasses, text books, and paperwork. A professional example is searching for tumors in a computed tomography scan; the target tumor must be discerned from distractors such as cysts, tissue, and muscle. Successful visual search requires the deployment of attention, followed by target detection, recognition, and identification (i.e., that a target is present and the type of target can be discerned).
A popular visual search theory, Feature Integration Theory (FIT), by Treisman and Gelade (1980) (as cited in Beck, Lohrenz, & Trafton, 2008) proposed that basic features are identified in parallel and then selective attention binds these features into objects serially. FIT assumes that a single basic visual feature (e.g., color or shape) distinguishes a target from distractors (Pashler, 1998; as cited in Koller, Drury, & Schwaninger, 2009), that search is exhaustive and objects are never checked twice, and that objects are processed one at a time (Chun & Wolfe, 1996; as cited in Pashler, 1998). Newer theories based on reaction time (i.e., search time) and set size (i.e., search size) counter the dichotomous division between serial and parallel search, the basis of FIT.

Chun and Wolfe (1996) (as cited in Pashler, 1998) proposed that search proceeds a target list until a target is found or until no items remain with activations that are above an "activation" threshold. The remaining items are deemed unlikely to be targets and are not visited by serial attention. In addition to this threshold mechanism, Chun and Wolfe proposed that some trials are terminated by guesses and that the probability of guessing increases as search time increases.

The newer theories also assume that multiple basic visual features distinguish a target from distractors (Pashler, 1998).

Visual search starts with the visual scanning of an area to be searched (Koller, Drury, and Schwaninger, 2009); during this process attention is directed to an object with the highest priority (Pashler, 1998). Target detection occurs next, here a potential target is identified. Target recognition occurs when a potential target matches a representation stored in visual memory. Visual search concludes by either deciding to stop searching or directing attention to a different potential target (Koller, Drury, and Schwaninger, 2009). If a potential target is rejected, attention will continually process subsequent potential targets (Wolfe, 1994; as cited in Pashler,
1998). Visual search occurs in order of similarity to the actual target (Duncan & Humphreys, 1989; as cited in Beck, Lohrenz, & Trafton, 2008). This process, while lengthy in description, is conducted very efficiently by humans. Koch and Ullman (1985) (as cited in Parasuramana, Greenwooda, & Alexander, 2000), found that “When search is distractor-dependent, objects (e.g., letters) can be searched at a rate of about 30-40 ms, which is consistent with the temporal characteristics of a covert attention mechanism” (p. 2). Factors that negatively impact search efficiency include distractor similarity, spatial layout, object occlusion, and background complexity (Beck, Lohrenz, & Trafton, 2008). Think about security screening operators searching air travelers’ property for Improvised Explosive Devices (IEDs). A pen or tie clip may appear very similar to a detonator, a primary IED component; it may not be oriented canonically (optimum spatial layout), and the traveler’s property may be heavily cluttered (object occlusion and background complexity).

Complex visual search, such as searching for tumors in computed tomography scans, is highly fatiguing due to high negative consequences, time stress, high and low workload periods, weak and infrequent targets, and high levels of background noise. This results in degraded performance with time. Since humans are generally not good at prolonged complex visual search tasks, Mosier and Skitka (2006) (as cited in as cited in Goh & Wiegmann, 2006) stated that automation could be used to mitigate human limitations. According to Goh and Wiegmann (2006), this can be accomplished through the use of decision support aids that “reduce the size of the visual search field and reduce the impact of distracters, shift attention to cued locations and engage attentional focus, and improve the sensitivity of observers” (p. 16). Humans can then scrutinize characteristics of potential targets within regions indicated by decision support aids, assess their nature, and make final determinations in a manner intended to be more effective and efficient.
LITERATURE SYNTHESIS

This thesis was founded on the basis that humans are highly social beings and that a virtual agent could extrapolate the concept of trust in people to trust in automation, and improve complex visual search task performance. When synthesized the reviewed literature topics (human behavior, human social trust, virtual representation of human embodiment, automation, automation reliability, automation reliance, automation trust, and visual search) support this thesis research. The following literature findings interrelate these topics:

- Humans respond socially to technology; human-human and human-computer interactions may be similar (Reeves & Nass, 1996; as cited in Lee & See, 2004).
- When knowledge is lacking on the risks and benefits of technology, people rely on social trust to make judgments (Siegrist & Cvetkovich, 2000).
- Virtual representations of human embodiment can provide human operators a more natural computer interface; operators are more likely to develop social relationships with them (Morrison, 2009).
- A function of automation is to present information that assists humans with problem solving and decision making. Decision support aids take information from the environment, integrate it with other information sources, and present a recommendation to humans (Adams, Bruyn, Houde, & Angelopoulos, 2003).
- When automation is reliable, human performance improves (e.g., less errors) (Skitka, Mosier, & Burdick, 1999; as cited in Cummings, 2004).
- Madhavan, Wiegmann, and Lacson (2003) (as cited in Goh & Wiegmann, 2006) found that when automation makes “easy” mistakes, as compared to “hard” mistakes, reliance is affected more.
• Automation reliance is necessary for trust in automation to grow (Muir & Moray, 1996; as cited in Goh & Wiegmann, 2006).
III. METHOD

This thesis research was designed to examine performance and automation trust with imperfect decision support aids in a domain of high complexity and high consequence. The assessment was conducted with a traditional decision support aid (bounding box) and a virtual agent, presented as a human-like graphical head with low-guidance verbal prompts.

PARTICIPANTS

Professional image analysts (e.g., radiologists, security screening operators, air traffic controllers) did not participate in this research study, rather a convenience sample of 36\(^1\) participants ranging in age from 18 to 63 participated (refer to Figure 1). Use of a convenience sample is satisfactory according to Gallway and Drury (1986) (as cited in Khasawneh, Kaewkuekool, Bowling, Desai, Jiang, Duchowski, & Gramopadhye, 2003), who showed that nominal differences exist between actual inspectors employed to perform visual search tasks and non-actual inspectors.

\[^{1}\] An a priori power analysis was conducted using the software package G*Power (Version 3.1.9.2) to determine the minimum number of participants required. A total of 36 participants were included in the data analysis, with an effect size of 0.5, \(\alpha\)-level of 0.05, and 1-\(\beta\) of 0.73.
Figure 1. Participants Age. This figure illustrates the age of the participants; the mean age was approximately 36 years.

All participants were screened to normal 20/20 Snellen visual acuity or better, with or without optical correction. Additionally, all participants completed high school and (at least) some college coursework (refer to Figure 2).

Figure 2. Highest Level of Education Completed. This figure illustrates participants’ educational backgrounds.
MATERIALS

Pre-Test Survey

Participants completed a pre-test survey on general demographics and their attitude toward technology. The demographic subsets included gender, age, and highest level of education completed. An adapted version of the Technology Attitude Survey (TAS) (refer to Appendix B) was used to determine participants’ attitude toward technology. The TAS was subjected to a validation study; a Cronbach alpha reliability of 0.92 was measured and results indicated that a single dimension explained item intercorrelations (McFarlane, Hoffman, & Green, 1997).

Vision Instrument

A Snellen Chart was used to measure each participant’s visual acuity from a distance of 20 feet.

Stimuli

Due to information sensitivity, real-world complex visual search task images (e.g., baggage X-rays, medial X-rays, satellite imagery) were not used as stimuli. Rather images comprised of alphabetic characters were developed; each image was based on a 2025-cell grid (45 cells X 45 cells). These alphabetic character images were considered high fidelity images (i.e., representative of a complex visual search task which requires the detection of a known target from a more complex array) (refer to Figure 3).
Figure 3. Alphabetic Character Image. This figure illustrates the experimental stimuli.

Each image contained uppercase letters (A, K, M, N, W, X, Y, or Z) that were randomly assigned (per grid cell) and equally distributed. “Target” images contained one of the following uppercase letters: C, G, O, Q, or U. **Note:** The background letters were drawn with only straight lines; the target letters were drawn with only curved lines or a combination of straight and curved lines. Each image was randomly generated in Microsoft Excel 2007 based on a nested if-then-else routine.

The “Spatial Cue” and “Spatial Cue + Virtual Agent” conditions included a bounding box to help participants determine the presence or absence of a target. A bounded region consisted of 400 cells outlined by a red un-filled rectangular bounding box; the outline weight was ¾ points. Consistent dimensions were used as bounding box precision was not an independent variable of interest. For “Target” images the center of the bounding box, in relation to the target, was randomized.

Virtual Agent

SitePal, a dynamic 3-D character building software, was used to develop the virtual agent. The virtual agent was an adult Caucasian male with its head, neck, and upper shoulders
displayed. To avoid the conveyance of specific behaviors (e.g., lack of interest, boredom) the virtual agent’s emotion and posture were neutral, and head nods, posture, and gaze shifts were randomly timed. While emotion, facial expression, posture, and gesture are important virtual agent characteristics, Durlach and Slater’s (1998) (as cited in Fabri, Moore, & Hobbs, 2002) research found that “avatars with rather primitive expressive abilities may engender strong emotional responses in people using a Collaborative Virtual Environment (CVE) system” (p. 2). The virtual agent conveyed low-guidance verbal prompts (“There may be a target here”, “I think there's a target”, “Is that a target?”, “Search for a target”, or “Look over there”) that were repeated twice, and randomly assigned to each image.

Test Platform

Custom software prepared in Microsoft Visual Studio .NET C#, presented in a fixed-viewing position on an external 24” high-resolution (1920 x 1080) color monitor, and controlled by a Lenovo Yoga 2 Pro laptop computer was used. Additionally, a standard mouse was used. The software provided a test platform that displayed stimuli to participants, and collected their accuracy and response time measurements. The test platform was comprised of six features (refer to Figure 4).

1. Stimuli Window – An image consisting of alphabetic characters.

2. Virtual Agent Window – For the “Spatial Cue + Virtual Agent” condition, a virtual agent was displayed in this window.

3. Timer – For each trial, a countdown timer was displayed on-screen to visually inform participants of the time remaining until a decision was required. The six second countdown timer conjured the fast-paced nature of complex visual search tasks. When time elapsed, the Stimuli Window was blacked out and participants were prompted to
select a decision button (“Target Present” or “Target Absent”) if one was not provided prior to the timer end.

4. Decision Radio Buttons – Participants were instructed to select the “Target Present” button if a target was present; the “Target Absent” if a target was not present. Upon selecting a button, the Stimuli Window was blacked out, unless it previously blacked out due to elapsed time.

5. Confidence Rating Radio Buttons – Participants were instructed to select one radio button to indicate their level of agreement (1 indicates Not Confident At All and 5 indicates Very Confident) with the following statement: I am confident with my decision.

6. Next Image – This button was activated for each trial subsequent to the selection of one Decision Radio Button and one Confidence Rating Radio Button. Upon selecting this button, the next alphabetic character stimuli image was displayed in the Stimuli Window.

*Figure 4. Test Platform. This figure illustrates the test platform that participants interacted with.*
Post-Test Survey

Upon completing all trials in their respective condition, participants completed a post-test survey on automation trust. The Empirically Derived (ED) scale (refer to Appendix C) was used; this scale was developed to address abstract trust in automation (i.e., without reference to an actual system). It was subjected to a validation study and has been used by Master, Jiang, Khasawneh, Bowling, Grimes, Gramopadhye, & Melloy (2005) who conducted research on trust over time in hybrid inspection systems (Chien, Semnani-Azad, Lewis, & Sycara, 2014).

Signal Detection Theory

Parasuraman, Masalonis, and Hancock (2000) stated that “Signal Detection Theory (SDT) could arguably be viewed as one of the most robust and useful quantitative theories in psychology” (p. 19). Combined with Jiang, Srinivasan, Gramopadhye, and Ferrell’s (2002) (as cited in Jiang, Khasawneh, Kaewkuekool, Bowling, Melloy, & Gramopadhye, 2003) statement that SDT is commonly used to model the decision making process in an inspection task, SDT provides a human decision-making behavior framework for visual search tasks.

The SDT framework (refer to Table 3) is founded on four outcomes (Hit, Miss, False Alarm, and Correct Rejection) that result from characteristics of a signal, and a human’s physical and psychological state. Characteristics of a signal, both environmentally and neurally, are always embedded in ‘noise’ or random variation. ‘Noise’ can range from acoustic noise (e.g., high dBA) to variations in human state (e.g., confident one minute and unconfident the next, greater fatigue in the evening versus the morning). For all intents and purposes, ‘noise’ can be any property that decreases the saliency of a signal (Szalma & Hancock, n.d.).

<table>
<thead>
<tr>
<th>Target</th>
<th>Respond “Target Absent”</th>
<th>Respond “Target Present”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miss</td>
<td>Hit</td>
<td></td>
</tr>
<tr>
<td>Correct Rejection</td>
<td>False Alarm</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Signal Detection Theory Measures
The four outcomes reflect several target detection performance measures. Probability of Detection is the proportion of “Target” trials to which a participant responds “Target Present” when the target is present (= P (“Target Present” | “Target”)) and Probability of False Alarm is the proportion of “Non-Target” trials to which a participant responds “Target Present” when the target is not present (= P (“Target Present” | “Non-Target”)).

Probability of Detection and Probability of False Alarm reflect two measures: sensitivity (commonly measured by d-prime) and response bias (commonly measured by criterion placement measure). Sensitivity and response bias recognize the distinction between human sensation, perception, and cognition, including measures of these indices (Szalma & Hancock, n.d.). These measures “paint the whole picture” as credit is assigned for hits and a penalty is assigned for false alarms. Sensitivity refers to how hard or easy it is to discriminate between target and noise items, and the overlap function between the two distributions; it was calculated by (= z(Probability of Detection) – z(Probability of False Alarm)). When targets are more similar to non-targets, signal and noise distributions move closer to one another, resulting in lower levels of sensitivity. Sorkin and Woods (1985) stated that d-prime values near zero will yield performance at chance levels and values above 4.0 will yield essentially errorless performance” (p. 5). Response bias, independent of sensitivity, refers to the extent in which one response is more probable than another; it was calculated by (= -0.5(z(Probability of Detection) + z(Probability of False Alarm))). A response bias value greater than zero indicates a bias toward responding “Target Absent”; a value less than zero indicates a bias toward responding “Target Present”; a value equal to zero indicates no bias.

Jiang, Khasawneh, Kaewkuekool, Bowling, Melloy, and Gramopadhye (2003) correlated trust with SDT measures in a hybrid inspection system. Both humans and computers searched for defects on printed circuit boards; final decisions were made by the humans (i.e., humans
were permitted to override decisions made by the computers. A SDT analysis indicated that the larger the response criterion, the more trust inspectors had in automation. The researchers attributed this to a more conservative system with fewer false alarms. Koller, Drury, and Schwaninger (2009) correlated the effect of training and demographics with SDT measures in an airport security screening task. A between-subjects design was employed; one group of participants received additional training on recognizing weapons in X-ray images of bags, while another group did not. The effect of training increased the Probability of Detection and decreased the Probability of False Alarm. Significant correlations were also found between sensitivity and age, and sensitivity and years on the job (by threat category).

DESIGN

The experiment employed a 3 (Condition: “Control”, “Spatial Cue”, “Spatial Cue + Virtual Agent”) x 2 (Difficulty Level: “Easy” and “Hard”) factorial design. Participants were divided into three groups so that there were an equal number of participants in each condition. Participants in the “Control” condition did not receive support from a decision support aid. Participants in the “Spatial Cue” condition received support from a decision support aid in the form of a bounding box. Participants in the “Spatial Cue + Virtual Agent” condition received support from a decision support aid in the form of a bounding box and virtual agent (a human-like graphical head that provides low-guidance verbal prompts such as, “There may be a target here”).

Image difficulty (“Easy” or “Hard”) was based on clutter; clutter was manipulated by varying the white space (blank cells) within an image. “Easy” images were defined as 20 percent clutter; “Difficult” images as 30 percent clutter. Images were classified as “Target” or “Non-Target”. Similar to professional complex visual search tasks, where image analysts know their targets (e.g., radiologist searching for a tumor, security screeners searching for weapons), participants were informed of targets (letters C, G, O, Q, or U) to search for. This provided
participants with knowledge to guide their visual search task. “Target” images included one target letter and “Non-Target” images did not include a target letter; the ratio of “Target” to “Non-Target” images was 1:3.

The decision support aids’ reliability was held constant at one level, 70 percent, which reflected error rates of 0.20 and 0.10 for false alarms and misses, respectively. Participants were not informed as to the actual reliability of the aid before testing.

Each participant (N = 36) performed 192 visual search trials for their assigned condition; 24 “Easy Target” trials, 24 “Hard Target” trials and 144 “Non-Target” trials. Participants received the same set of stimuli, but with different, random and counterbalanced orders. Hit and false alarm rates were identical for each condition.

PROCEDURE

The experiment was comprised of four main sections: Preliminary Tasks, Training, Assessment, and Concluding Tasks. The four sections did not exceed one hour and breaks were provided between each section.

Preliminary Tasks

Upon arrival, each participant was introduced to the researcher and completed a visual acuity test. The researcher then reviewed the informed consent form (refer to Appendix A), which described the general purpose of the study and experimental procedure; participants were then asked to provide informed consent by signing the form. Participants then completed a pre-test survey (refer to Appendix B).

Training

Seated in front of a workstation, the researcher trained each participant on the test platform. Each Graphical User Interface (GUI) element (e.g., window pane, button) was
described along with its functionality in relation to the experiment. Participants then received training on the targets according to the following instructions:

“Please search each image for the following targets (letters C, G, O, Q, or U). Some images will include a target, others will not. If an image contains a target, only one will be present.”

An informal exercise was then conducted to confirm participants’ comprehension of the targets (refer to Figure 5). Note: The questions were displayed one at a time; the “What are the target letters?” question was displayed four times (alternated with the other questions). Participants were required to pass (100 percent) the Target Comprehension Check in order to continue in the study.

![Figure 5. Target Comprehension Check. This figure illustrates the Target Comprehension Check that was used to verify participants’ comprehension of the targets.](image)

Participants then completed a training/exploration session in an individual self-paced manner and were allowed to ask questions. Participants received support from a decision support aid in the form of a bounding box for the training/exploration session. The decision
support aid was 100 percent reliable during the training/exploration session to establish trust and reliance.

Assessment

Participants then started their assigned experimental condition. Participants were asked to imagine that they were performing a complex visual search task according to the following instructions:

“Please imagine that you are completing a time-sensitive task. For example, imagine yourself as an assembly line worker at an automotive plant; your job is to inspect vehicle parts for manufacturing defects. You must complete this task quickly and only stop the production line if a defect is present. While your first concern is to identify defects which will increase drivers’ safety, please remember that it is extremely costly and time intensive to stop the production line. Similarly, for this experiment, your job is to search for a target (letters C, G, O, Q, or U). Remember, some images will include a target (only one of the target letters will be present), others will not. You should render your decision as soon as possible by selecting a decision button; you do not have to wait for the countdown timer to elapse. However, if you let the time completely elapse the image will be removed from the screen and you will be asked to render a decision.

At times, a decision support aid may assist you. The aid is intended to notify you that a target is present. The aid is not perfect and may be incorrect at times. It may notify you that a target is present when one is not. Also, it may “forget” to notify you when a target is present. It is important to search the entire image. After rendering your decision, please select your confidence level [with your decision] and select “Next Image”. Please repeat this process for the remaining images.”
Concluding Tasks

Upon completing all trials in their respective condition, participants completed a post-test survey on automation trust. Participants were then provided with instructions on how to obtain further information about the study and thanked for their time.
IV. RESULTS

The following analyses were conducted to identify important implications for the presentation of decision support aids, and more specifically to quantify any benefits imparted by a decision support aid in the form of a virtual agent. A summary of the results is presented in Figure 12.

DATA ANALYSIS

Data were analyzed using JMP (Version 11.0.0) and Microsoft Excel with the Analysis ToolPak (2013). The data analysis employed various statistical methods depending on the scientific question and dataset under consideration. Descriptive statistics were used to explain central values and variability within datasets. Three-way Analysis of Variance (ANOVA) with two levels of difficulty ("Easy" and "Hard") and three levels of decision aid presentation ("Control", "Spatial Cue", and "Spatial Cue + Virtual Agent") were used to analyze performance measures. Although explicit interaction hypotheses were not formulated, significant interactions were examined. A conventional level of \( p \leq .05 \) was used to determine statistically significant differences. For significant ANOVA results, post hoc analysis (Student’s t test) were conducted on all possible pairwise contrasts. In addition, raw effect sizes (denoted by \( \delta \)), as reported by the Power Analysis interface in JMP, were examined.

The technology attitude survey was analyzed as a 16-item Likert scale that was assessed with seven response categories ranging from “Very Untrue” to “Very True” (refer to Appendix B). Similarly, the trust survey was analyzed as a 12-item Likert scale that was assessed with seven response categories ranging from “Strongly Disagree” to “Strongly Agree” (refer to Appendix C).
The confidence survey was analyzed as Likert items; the scales consisted of four response categories ranging from “Strongly Disagree” to “Strongly Agree”.

TECHNOLOGY ATTITUDE

A composite score reflecting each participant’s attitude toward technology was calculated by summing the individual Likert item (i.e., sixteen criteria) responses per participant. The participants’ composite scores were then analyzed (means) per condition. Data showed a positive attitude toward technology among the three conditions: “Control” ($M = 90.67$, $SD = 14.78$), "Spatial Cue" ($M = 94.17$, $SD = 12.35$), and "Spatial Cue + Virtual Agent" ($M = 95.08$, $SD = 13.28$) (refer to Table 4).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean (M)</th>
<th>Standard Deviation (SD)</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>90.67</td>
<td>14.78</td>
<td>68</td>
<td>112</td>
</tr>
<tr>
<td>Spatial Cue</td>
<td>94.17</td>
<td>12.35</td>
<td>77</td>
<td>112</td>
</tr>
<tr>
<td>Spatial Cue + Virtual Agent</td>
<td>95.08</td>
<td>13.28</td>
<td>77</td>
<td>111</td>
</tr>
</tbody>
</table>

Note. A maximum score of 112 reflects the most positive attitude toward technology.

Participants’ attitude toward technology did not differ significantly between the three conditions, $F(2, 33) = 153.63$, $p > 0.35$, $\delta = 1.90$, as reported by a one-way ANOVA.

OVERALL PROBABILITY OF DETECTION

The Probability of Detection ($P_d$) was calculated for each condition by dividing the total number of correctly identified “Target Present” images by the total number of target images. A two-way ANOVA yielded a main effect for decision aid presentation, $F(2, 66) = 132.61$, $p < .0001$, $\delta = 0.18$. Post-hoc analysis indicated that detection was significantly higher for the “Spatial Cue + Virtual Agent” condition ($M = 0.84$, $SD = 0.06$) as compared to the “Spatial Cue” condition ($M = 0.77$, $SD = 0.11$) and the “Control” condition ($M = 0.43$, $SD = 0.12$). Additionally detection was significantly higher for the “Spatial Cue” condition as compared to the “Control” condition. The main effect of difficulty yielded a test statistic of $F(1, 66) = 10.45$, $p = .001$, $\delta = 0.04$; detection
was significantly greater for “Easy” images ($M = 0.71, SD = 0.19$) as compared to “Hard” images ($M = 0.64, SD = 0.22$). The interaction effect was not significant, $F(2, 66) = 1.17, p = .16$ (refer to Figure 6).

**Figure 6.** Probability of Detection. This figure illustrates the main effect of decision aid presentation (a) and difficulty (b), and the non-significant interaction of both on detection performance (c).

**OVERALL PROBABILITY OF FALSE ALARM**

Probability of False Alarm ($P_{fa}$) was calculated by dividing the total number of incorrectly identified “Target Present” images by the total number of non-target images. A two-way ANOVA yielded a main effect for decision aid presentation, $F(2, 66) = 10.32, p < .0001, \delta = 0.03$. While the false alarm rate was lower for the “Spatial Cue + Virtual Agent” condition ($M = 0.07, SD = 0.05$) as compared to the “Spatial Cue” condition ($M = 0.09, SD = 0.06$), this difference was not significant as indicated by a post-hoc analysis. The “Control” condition ($M = 0.03, SD = 0.03$) was significantly lower than the other two conditions. The main effect of difficulty yielded a test statistic of $F(1, 66) = 7.11, p = .005, \delta = 0.01$; false alarm performance was significantly lower for
“Easy” images ($M = 0.05$, $SD = 0.05$) as compared to “Hard” images ($M = 0.08$, $SD = 0.06$). The interaction effect was not significant, $F(2, 66) = 0.46$, $p = .0.32$ (refer to Figure 7).

(a) (b) (c)

Figure 7. Probability of False Alarm. This figure illustrates the main effect of decision aid presentation (a) and difficulty (b), and the non-significant interaction of both on false alarm performance (c).

OVERALL SENSITIVITY

Task sensitivity was analyzed by deriving d-prime ($d'$) for each condition; the difference between the z-transforms of $P_d$ and $P_{fa}$ was calculated. A two-way ANOVA yielded a main effect for decision aid presentation, $F(2, 66) = 21.54$, $p < .0001$, $\delta = 0.36$. Post-hoc analysis indicated that sensitivity was significantly higher for the “Spatial Cue + Virtual Agent” condition ($M = 2.60$, $SD = 0.43$) as compared to the “Spatial Cue” condition ($M = 2.21$, $SD = 0.57$) and the “Control” condition ($M = 1.76$, $SD = 0.52$). Additionally sensitivity was significantly higher for the “Spatial Cue” condition as compared to the “Control” condition. The main effect of difficulty yielded a test statistic of $F(1, 66) = 23.34$, $p < .0001$, $\delta = 0.25$; sensitivity was significantly greater for
“Easy” images ($M = 2.44, SD = 0.50$) as compared to “Hard” images ($M = 1.93, SD = 0.61$). The interaction effect was not significant, $F(2, 66) = 0.29, p = .38$ (refer to Figure 8).

![Figure 8](image)

**Figure 8.** Overall Sensitivity. This figure illustrates the main effect of decision aid presentation (a) and difficulty (b), and the non-significant interaction of both on task sensitivity (c).

OVERALL RESPONSE BIAS

Response bias was analyzed by deriving the criterion placement measure (c) for each condition; the sum of the z-transforms of $P_a$ and $P_{1a}$ at negative half value (-0.5) was calculated. A two-way ANOVA yielded a main effect for decision aid presentation, $F(2, 66) = 71.84, p < .0001, \delta = 1.15$. A liberal response bias (i.e., a tendency to say "Target Present" more than "Target Absent") was found for both the “Spatial Cue + Virtual Agent” condition ($M = -2.08, SD = 0.37$) and “Spatial Cue” condition ($M = -1.82, SD = 0.47$); a post-hoc analysis indicated that the difference between the two conditions was not significant. The “Control” condition attained the lowest amount of response bias ($M = 0.49, SD = 1.30$) which was significantly different than the than the other two conditions. It is important to note that this bias was conservative (i.e., a tendency to say "Target Absent" more than "Target Present") as compared to the other
conditions. The main effect of difficulty yielded a test statistic of $F(1, 66) = 4.48, p = .019, \delta = 0.2$; response bias (liberal) was significantly greater for “Easy” images ($M = -1.34, SD = 1.44$) as compared to “Hard” images ($M = -0.93, SD = 1.38$). The interaction effect was not significant, $F(2, 66) = 0.09, p = .0.46$ (refer to Figure 9).

![Figure 9](image)

**Figure 9.** Overall Response Bias. This figure illustrates the main effect of decision aid presentation (a) and difficulty (b), and the non-significant interaction of both on response bias (c).

**OVERALL RESPONSE TIME**

Response Time (RT) was calculated by subtracting $T_{Start}$ (image is displayed) from $T_{End}$ (a decision is rendered or time elapsed). A two-way ANOVA yielded a main effect for decision aid presentation, $F(2, 66) = 85.98, p < .0001, \delta = 0.57$. Post-hoc analysis indicated that RT was significantly faster for the “Spatial Cue + Virtual Agent” condition ($M = 5.20, SD = 0.17$) as compared to the “Control” condition ($M = 5.68, SD = 0.11$), however the “Spatial Cue” condition achieved the fastest RT ($M = 4.31, SD = 0.59$). Image difficulty did not produce a main effect on participants' RT (“Easy” images ($M = 5.05, SD = 0.65$) and “Hard” images ($M = 5.08, SD = 0.70$)).
$F(1, 66) = 0.15, p = 0.35$ or an interaction with decision aid presentation, $F(1, 66) = 0.31, p = 0.37$ (refer to Figure 10).

![Figure 10](image)

Figure 10. Overall Response Time. This figure illustrates the main effect of decision aid presentation (a), and the non-significant effect of difficulty (b) and the interaction of both on response time.

TRUST

A composite score reflecting each participant’s system-level trust was calculated by summing the individual Likert item (i.e., twelve criteria) responses per participant. The participants’ composite scores were then analyzed (means) per condition. Data showed an increase in trust between the three conditions: “Control” ($M = 51.83, SD = 8.98$), “Spatial Cue” ($M = 59.00, SD = 8.64$), and “Spatial Cue + Virtual Agent” ($M = 67.75, SD = 7.45$) (refer to Table 5).

<table>
<thead>
<tr>
<th>Table 5 Trust Survey Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
</tr>
<tr>
<td>Control</td>
</tr>
<tr>
<td>Spatial Cue</td>
</tr>
<tr>
<td>Spatial Cue + Virtual Agent</td>
</tr>
</tbody>
</table>
Trust differed significantly between the three conditions, $F(2,33) = 10.847$, $p = 0.0002$, $\delta = 6.51$, as reported by a one-way ANOVA. Post-hoc analysis indicated that trust was significantly higher for the “Spatial Cue + Virtual Agent” condition ($M = 67.75$, $SD = 7.45$) as compared to the “Spatial Cue” condition ($M = 59.00$, $SD = 8.64$) and the “Control” condition ($M = 51.83$, $SD = 8.98$). Additionally trust was significantly higher for the “Spatial Cue” condition as compared to the “Control” condition.

**CONFIDENCE**

Participants’ confidence was determined by tabulating (refer to Table 6) their level of agreement ("Strongly Disagree", "Disagree", "Agree", and "Strongly Agree") to the following statement\(^2\): “I am confident with my decision.”

<table>
<thead>
<tr>
<th>Condition</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>653 (28.34%)</td>
<td>272 (11.81%)</td>
<td>719 (31.21%)</td>
<td>660 (28.65%)</td>
</tr>
<tr>
<td>Spatial Cue</td>
<td>3 (0.13%)</td>
<td>405 (17.58%)</td>
<td>1464 (63.54%)</td>
<td>432 (18.75%)</td>
</tr>
<tr>
<td>Spatial Cue + Virtual Agent</td>
<td>66 (2.86%)</td>
<td>777 (33.72%)</td>
<td>633 (27.47%)</td>
<td>828 (35.94%)</td>
</tr>
</tbody>
</table>

Participants in the "Spatial Cue + Virtual Agent" condition were the most confident, with the highest selection of "Strongly Agree" (35.4 percent), as compared to the "Spatial Cue" (18.75 percent) and "Control" (28.65 percent) conditions. It is important to note that low confidence levels were the least salient in the "Spatial Cue" condition (0.13 percent), as compared to the "Control" (28.34 percent) and "Spatial Cue + Virtual Agent" (2.86 percent) conditions. Additionally, the response frequencies across conditions for each image type (difficulty) were similar (refer to Figure 11).

\(^2\) Single Likert-item.
RESULTS HIGHLIGHTS

The results highlights (refer to Figure 12) the fact that a novel decision support aid in the form of a "Spatial Cue + Virtual Agent" offers increased benefits for a complex visual search task.

**Figure 12. Results Highlights.** This figure highlights results from this thesis research.
V. DISCUSSION

To improve users’ performance on complex visual search tasks, technology is commonly outfitted with automation that displays traditional spatial cues (e.g., bounding box, coloration) to support users’ decision making. While traditional spatial cues afford users many benefits, as compared to the absence of a decision support aid, previously cited literature indicates concern surrounding the adequacy of spatial cues. Spatial cues do little to convey information on the inner workings of automated processes, creating a sense of obscurity for users. This in turn negatively impacts users’ trust, arguably the most important subjective construct effecting the appropriate use of automation.

This research investigated a possible solution, the integration of virtual agents in graphical user interfaces. By increasing the humanness of a decision support aid, improvements were found for a complex visual search task. These findings corroborate research conducted by Foster (2007) (as cited in Kuhnel, Weiss, & Moller, 2009) in that an Embodied Conversational Agent (ECA) enhanced general human-computer interaction and improved user satisfaction. Additionally, these improvements were found without varying the reliability of the decision support aids investigated, demonstrating that the presentation of information alone can improve performance. These findings may assist in the design of graphical user interfaces for complex systems, influence the deployment of automation (including software updates), and modify training approaches.
The primary objective of this study was to evaluate the effects of a virtual agent on human performance in a complex target detection task. This was accomplished by comparing users performance with a traditional decision support aid ("Spatial Cue") against performance with a novel decision support aid ("Spatial Cue + Virtual Agent"), both with the same reliability. **Note:** A decision support aid was not displayed for the positive control group.

Hypotheses of this research naturally define improved human performance as improved task sensitivity (i.e., the ability to discriminate between target and non-targets), increased [appropriate] reliance on a decision support aid, or a combination of both. An examination of the d-prime and criterion placement measure (c) data support these hypotheses, and as such improved performance with a "Spatial Cue + Virtual Agent". The d-prime results were as expected; accuracy increased from the unaided condition ("Control") to the aided (standard) "Spatial Cue" condition to the aided (novel) "Spatial Cue + Virtual Agent" condition. These results support previously cited literature in that decision support aids improve accuracy by guiding users’ attention to potential target regions. However, a new finding emerged: a significant increase in accuracy with a “Spatial Cue + Virtual Agent”. The added virtual agent likely increased the overall salience of the standard aid, increasing accuracy. In other words, the virtual agent may have provided users with more meaningful information to inform their decision-making, whereas additional information was not communicated by the “Spatial Cue” alone. As previously stated, improved human performance is also defined as increased [appropriate] reliance on a decision support aid. Automation reliance was investigated by examining the difference in criterion placement measure (c) results between the three conditions. Participants’ liberal bias increased between each of the three conditions, indicating continued reliance on the aid. Overall, as hypothesized, participants demonstrated an
appropriate reliance on the "Spatial Cue + Virtual Agent" decision support aid, which in turn improved their accuracy.

To further investigate the impact of the "Spatial Cue + Virtual Agent", the performance measures comprising users’ sensitivity and response bias were analyzed separately. The Probability of Detection results were as expected, with a 34 percent increase found between the unaided condition ("Control") to the aided (standard) "Spatial Cue” condition, and an added seven percent increase found with the "Spatial Cue + Virtual Agent". The Probability of False Alarm results were unexpected, and suggested that no aid would optimize false alarm performance. This contradicts the basic theme (aided performance is superior to unaided performance) of this research. One possibility for this finding may reside with the experimental stimuli; the targets may have been easily distinguished from the distractor items. An interesting finding, while not statistically significant, was that participants in the "Spatial Cue + Virtual Agent" condition were more inclined to override the automation, resulting in lower false alarms than the "Spatial Cue" condition. Otherwise stated, participants were more likely to correctly disagree with a “human” rather than a computer. The results also indicate that the "Spatial Cue + Virtual Agent" did not produce adverse effects on false alarm performance.

EFFICIENCY

The "Spatial Cue” condition achieved the fastest response time, thereby failing to support the hypothesis that response time would decrease with a "Spatial Cue + Virtual Agent" (as compared to the "Spatial Cue"). This suggests that the "Spatial Cue + Virtual Agent" increased visual search time, however it is reasonable to assume that the novelty of the virtual agent may have caused this effect, and possibly distracted participants from the visual search task. This, along with the simulated time stress may have produced this effect. Increased use with virtual agents will likely mitigate this impact and produce the hypothesized effect. It is
important to note that the difference between response times was less than one second; a nominal difference offset by the fact that participants were more accurate, as indicated by the sensitivity data, in their decision-making when a “Spatial Cue + Virtual Agent” was present.

SUBJECTIVE MEASURES

Participants’ attitude toward technology, measured pre-test, did not differ significantly between the three conditions. Each condition averaged 90 points (out of a possible 112 points) or greater, suggesting a positive view of technology. Other survey constructs point to a high interest, need, and use of technology, and ability to learn technology. This similar state of mind supports several important aspects of the research methodology, from the administration method (computer-based) to the random assignment of participants. Confounding demographic variables (e.g., gender, age, highest level of education completed) may have been reduced between the three conditions. Additionally, participants’ similar positive attitude toward technology ‘in general’ (i.e., without consideration to automated technology) was important since the experiment introduced varying forms of imperfect automation per condition.

Participants’ level of confidence, measured post-decision during the test, was the greatest with the "Spatial Cue + Virtual Agent". Participants were more successful at discerning targets from non-targets with the "Spatial Cue + Virtual Agent", which could be attributed to increased confidence in their decision making capabilities. The d-prime and criterion placement measure (c) results indicate that this was achieved without causing overconfidence or self-reliance (negative effects of increased confidence).

Participants’ level of trust, measured post-test, was significantly greater when the "Spatial Cue + Virtual Agent" was present, compared to the other decision support aids. Based on participants’ increased confidence with the "Spatial Cue + Virtual Agent" and in accordance with literature from Madsen and Gregor (2000) (as cited in Adams, Bruyn, Houde, &
Angelopoulos, 2003) who defined trust in a decision aid as, “...the extent to which a user is confident in, and willing to act on the basis of the recommendations, actions, and decisions of an artificially intelligent agent”, this finding was anticipated.

LIMITATIONS

The results of this study provide a useful “first look” on the effects of virtual agents for complex visual search tasks, however several limitations impact the conclusions derived. First, images comprised of alphabetic characters do not have the same fidelity as real-world images. Undoubtedly more demanding visual search tasks exist (e.g., security screening, medical screening) and are important to investigate. Second, image manipulation tools (e.g., Zoom, Rotate) common to complex visual search tasks were not provided to participants. Third, the simulated stressors (e.g., response time limit, infrequent targets) could not fully represent a real screening environment with real consequences; as such the true effects of workload, stress, and vigilance were not sufficiently investigated. Finally, professional screeners did not serve as participants. While previously cited research supported this part of the methodology, authentic end-users possess the most relevant skills, experience, task knowledge, and training due to conducting operations on a day-to-day basis in target environments. Nonetheless, the results demonstrate that a "Spatial Cue + Virtual Agent” improves performance with imperfect automation and that humans respond differently, albeit positively, to graphical user interfaces integrated with more human-like features.

CONTRIBUTIONS

Traditional decision support aids are commonly used to support complex visual search tasks. Thus, strategies to improve the utilization of automation have been limited to directly improving these traditional aids through coloration, segmentation, pixilation, etc., techniques. On a separate but related note, over much of the last decade, industry has primarily employed
virtual agents, avatars, embodied conversational agents, smart virtual assistants, etc., in technical applications of low complexity and low consequence. This research contributed to a changing view of these two applications by using a virtual agent to support a complex visual search task. This research demonstrated that automation utilization could be improved by increasing the humanness of automation; it differed from the commonly held view that improvements in automation utilization required increased automation reliability. The importance of this finding lies in the fact that significant advances in automation reliability, while undoubtedly important, usually require a considerable length of time. Now, virtual agents could mitigate negative effects of imperfect automation until advances in automation reliability are realized. Simply put, this research identified a new opportunity for engendering trust with imperfect automation.

FOLLOW-ON RESEARCH

The results presented herein are anticipated to support the need for additional research on the integration of virtual agents and complex visual search tasks. At the most basic level, current research limitations can be addressed by repeating this study in a more realistic setting with genuine stimuli and appropriate end-users.

Future research may also examine how the use of a "Spatial Cue + Virtual Agent" decision support aid evolves over time. It is hypothesized that many effects from this study could be strengthened with increased use/practice with virtual agents. Or a point of diminishing returns with regard to virtual agent exposure could be identified. Such findings could influence user training; perhaps it is only beneficial to expose novices to virtual agents in efforts to mitigate immediate bias with automation. Or perhaps it is only beneficial to expose users to virtual agents when automation is first deployed or when automation reliability is low.
This study held the reliability level constant, so the functional relationship between reliability and performance could not be determined. Future research could parametrically manipulate reliability across several discrete levels and users’ performance data could be examined for a point of diminishing returns. Similarly, only two levels of difficulty were assessed in this study, which may have contributed to the lack of interaction effects between image difficulty and the decision support aids. If more discrete levels of difficulty were assessed, a "Spatial Cue + Virtual Agent" decision support aid may only be warranted for images of medium and high difficulty. Likewise, a "Spatial Cue" may only be warranted for images of lower difficulty, or perhaps these images should not be automated at all. This could increase user satisfaction by limiting aid to justified/critical cases. Additionally, aiding users appropriately could result in greater operational efficiencies, as time must be taken to resolve automation output.

To observe the full possibilities of a "Spatial Cue + Virtual Agent", an enhanced virtual agent should be designed for future research. To further imitate human-to-human/face-to-face communication, other characteristics (e.g., emotion, mixed initiative, back channeling, sense of presence) could be incorporated in the design of virtual agents. Bickmore (2003) (as cited in Nickerson & Reilly, 2004) found that participants trusted a conversational agent more when it elicited and expressed affect, even though the participants did not believe the machine itself was experiencing emotion. At a more advanced level, virtual agent features, such as affect, could be varied based on stimuli characteristics. For example, a highly dense X-ray image may mask abnormalities in a medical scan, increasing the probability of a [automation] false alarm. In this case, virtual agents may convey high uncertainty through raised eyebrows, increased eye saccades, a creased forehead, or a naturally intense gaze.
This study utilized virtual agents to represent automation for a complex visual search task; the use of virtual agents to represent other humans in complex environments should also be investigated. Complex systems often involve multi-level decision-makers. Virtual agents could increase trust, collaboration, the sense of immediacy, etc., by adding humanness to remotely located individuals, or increase the credibility of earlier decisions or information used throughout the entire decision-making process.

The follow-on research recommendations described above will add authenticity to the machine (automation) and environment when investigating the effects of a virtual agent on complex visual search performance. These advances will likely elicit greater trust from users, resulting in performance improvements. That is, of course, if users have an inherent disposition to trust machine (automation) and for that matter, other humans. Merritt and Ilgen (2008) stated that little empirical attention has been extended to individual differences as a predictor of trust, as compared to automation characteristics. Future research should investigate human personality traits (e.g., agreeableness, flexibility, cooperation, open-mindedness) and their ability to predict users’ success with a virtual agent for a complex visual search task. Then, accounting for the entire system (human, machine, and environment), the most important benefits of integrating virtual agents in complex graphical user interfaces could be realized.

The results of this study will hopefully have wide-reaching benefits within fields conducting complex visual search tasks. The research findings indicate that a "Spatial Cue + Virtual Agent" decision support aid improves complex visual search performance. While some hypotheses were not supported, additional research is needed to produce more definitive results and maximize the use of virtual agents in complex environments.
VI. CONCLUSION

Decision support aids in complex visual search tasks offer advantages such as smaller visual search fields, reduced distracter effects, appropriate shifts in attention, engaged attentional focus, and improved sensitivity (Goh and Wiegmann, 2006). However, traditional decision support aids are obscure and cause trust issues with automation, which ultimately effects task performance. Nevertheless, this research generally supported the commonly held views on decision support aids, while also demonstrating that a novel decision support aid in the form of a "Spatial Cue + Virtual Agent" offers increased effectiveness for a complex visual search task. Additionally, the "Spatial Cue + Virtual Agent" was shown to likely have no more than a negligible impact on operational efficiencies. The results obtained from this study also support the following conclusions:

- General efficacy improvements were found with a "Spatial Cue + Virtual Agent" without requiring increased automation reliability
- A "Spatial Cue + Virtual Agent" significantly increased trust in automation
- A "Spatial Cue + Virtual Agent" improved decision-making confidence
- Appropriate reliance is achieved with a "Spatial Cue + Virtual Agent"
- The impact of image difficulty did not depend on the type of decision support aid
- Humans respond to virtual humans in a social manner

This study indicates that there is a potential to mitigate declines in automation trust as a consequence of obscure decision support aids by simply increasing aids’ humanness. Overall,
the integration of virtual agents for complex visual search tasks is a commendable goal for graphical user interface design.
REFERENCES


APPENDIX A: INFORMED CONSENT

WRIGHT STATE UNIVERSITY
INFORMED CONSENT FOR PARTICIPATION IN RESEARCH

Virtual Agent Interaction: Improving Cognitive Abilities and Trust for a Complex Visual Search Task

PLEASE READ THIS DOCUMENT CAREFULLY. Your signature is required for participation. This study is being conducted by Heather Milecki, a graduate student at Wright State University; the study has been reviewed by the Wright State University Institutional Review Board.

I. Participation Agreement

This consent form provides a description of the research study and outlines the risks, benefits, and confidentiality involved. You are entitled to ask questions about the study before deciding to participate. Your signature indicates that you have read the consent form, been informed of the content presented herein, and freely consent to participate in the research study. If you would like a copy of this consent form, please request one and it will be provided.

II. Research Purpose

The purpose of this research study is to evaluate automation aids and measures for a complex visual search task. More specifically, the effectiveness, efficiency, and acceptance of a standard automation aid (Spatial Cue) will be compared to a more novel aid (Spatial Cue + Virtual Agent).

III. Procedure

Participation in this research study is estimated to be approximately one hour and thirty minutes. You will begin by completing a demographic survey (e.g., gender, age, education) and technology attitude survey. You will then be asked to complete 120 visual search trials that vary in difficulty. You will be told which targets to search for and may receive assistant from an automation aid. For each trial, a six second countdown timer will be displayed on-screen to indicate time remaining. You will be prompted to select a decision button (“Target Present” or “Target Absent”) if you did not select a button prior to the timer end. A five minute break will be provided every 20 minutes to reduce the onset of fatigue. Upon completing all trials, you will complete a survey on automation trust.

IV. Risks

This research study involves no more than minimal risk to participants. Participants may feel fatigued or may experience blurred vision similar to what they would perceive while playing video games or working on a computer.

V. Benefits
There is no direct benefit to you as the result of participating in this study, however your participation will contribute to science and a better understanding of automation aids for a complex visual search task. The results of your participation may be leveraged to improve work conditions, equipment, and processes for complex visual search tasks in the future.

VI. Confidentiality and Anonymity

All data obtained are for research purposes only and will remain confidential. Research records will remain on a password protected computer or in locked file cabinets that will only be accessible for review by the researcher, the Institutional Review Board (IRB), and federal regulatory agencies. Participation is strictly anonymous and in no case will responses from your participation be identified. Rather, all data will be pooled and published in aggregate form only. Your name will only appear on this consent form; you will be assigned a participant number that will not be linked to this form or other research documents.

VII. Contact Information

Any technical questions or requests for further information about this research may be directed to:

Principal Investigator: Heather Milecki Phone: (937) 775-5044

The faculty advisor for this project is Dr. Jennie Gallimore who can be reached at (937) 775-4901.

Any questions regarding research participant rights may be directed to the Wright State University Institutional Review Board at (937) 775-4462.

VIII. Voluntary Consent

I have read this consent form and I volunteer to participate in this research study. I have been informed that I may decline to participate or withdraw from this research study at any time without penalty.

I have been informed that my consent does not take away any legal rights. I have been further informed that nothing in this consent form is intended to preempt any applicable federal, state, or local laws regarding informed consent.

Signature of Participant ___________________________ Date ___________________________

Printed Name of Participant ___________________________
APPENDIX B: PRE-TEST SURVEY

Virtual Agent Interaction: Improving Cognitive Abilities and Trust for a Complex Visual Search Task

PRE-TEST SURVEY

ID No

Gender

☐ Male
☐ Female

Age (Years)

Highest Level of Education Completed

☐ High school graduate, or equivalent
☐ Trade, technical, or vocational training
☐ Some college credit, no degree
☐ Associate degree
☐ Bachelor’s degree
☐ Master’s degree
☐ Doctorate degree

To what extent do you agree with the following statements?

1. Knowing how to use technology is a necessary skill for me.

☐ Very Untrue
☐ Untrue
☐ Somewhat Untrue
☐ Neutral
☐ Very True
☐ True
☐ Somewhat True

2. I like using technology.

☐ Very Untrue
☐ Untrue
☐ Somewhat Untrue
☐ Neutral
☐ Very True
☐ True
☐ Somewhat True

3. I feel confident with my ability to learn about technology.

☐ Very Untrue
☐ Untrue
☐ Somewhat Untrue
☐ Neutral
☐ Very True
☐ True
☐ Somewhat True
4. Working with technology makes me nervous.
   - Very Untrue
   - Untrue
   - Somewhat Untrue
   - Neutral
   - True
   - Somewhat True

5. I like using technology in my work.
   - Very Untrue
   - Untrue
   - Somewhat Untrue
   - Neutral
   - True
   - Somewhat True

6. I wish I could use technology more frequently.
   - Very Untrue
   - Untrue
   - Somewhat Untrue
   - Neutral
   - True
   - Somewhat True

7. Technology makes me feel stupid.
   - Very Untrue
   - Untrue
   - Somewhat Untrue
   - Neutral
   - True
   - Somewhat True

8. A job using technology would be very interesting.
   - Very Untrue
   - Untrue
   - Somewhat Untrue
   - Neutral
   - True
   - Somewhat True

9. I don't use technology much at work.
   - Very Untrue
   - Untrue
   - Somewhat Untrue
   - Neutral
   - True
   - Somewhat True

10. I'm not the type to do well with technology.
    - Very Untrue
    - Untrue
    - Somewhat Untrue
    - Neutral
    - True
    - Somewhat True

11. I feel uncomfortable using most technology.
    - Very Untrue
    - Untrue
    - Somewhat Untrue
    - Neutral
    - True
    - Somewhat True

12. Working with technology is boring.
    - Very Untrue
    - Untrue
    - Somewhat Untrue
    - Neutral
    - True
    - Somewhat True

13. I know that if I work hard to learn about technology, I will do well.
    - Very Untrue
    - Neutral
    - Very True
<table>
<thead>
<tr>
<th>Number</th>
<th>Question</th>
<th>Scale</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>I think using technology will be difficult for me.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Very Untrue</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Untrue</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Somewhat Untrue</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>True</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Somewhat True</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Technology makes me feel uneasy and confused.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Very Untrue</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Untrue</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Somewhat Untrue</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Very True</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>True</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Somewhat True</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Once I start using technology, I find it hard to stop.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Very Untrue</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Untrue</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Somewhat Untrue</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Very True</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>True</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Somewhat True</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please use the space below to provide additional information regarding the system.
APPENDIX C: POST-TEST SURVEY

Virtual Agent Interaction: Improving Cognitive Abilities and Trust for a Complex Visual Search Task

POST-TEST SURVEY

ID No. ______

To what extent do you agree with the following statements?

1. The system is deceptive.
   - [ ] Strongly Disagree
   - [ ] Disagree
   - [ ] Somewhat Disagree
   - [ ] Neither Agree or Disagree
   - [ ] Agree
   - [ ] Somewhat Agree

2. The system behaves in an underhanded manner.
   - [ ] Strongly Disagree
   - [ ] Disagree
   - [ ] Somewhat Disagree
   - [ ] Neither Agree or Disagree
   - [ ] Agree
   - [ ] Somewhat Agree

3. I am suspicious of the system’s intent, action, or outputs.
   - [ ] Strongly Disagree
   - [ ] Disagree
   - [ ] Somewhat Disagree
   - [ ] Neither Agree or Disagree
   - [ ] Agree
   - [ ] Somewhat Agree

4. I am wary of the system.
   - [ ] Strongly Disagree
   - [ ] Disagree
   - [ ] Somewhat Disagree
   - [ ] Neither Agree or Disagree
   - [ ] Agree
   - [ ] Somewhat Agree

5. The system’s actions will have a harmful or injurious outcome.
   - [ ] Strongly Disagree
   - [ ] Disagree
   - [ ] Somewhat Disagree
   - [ ] Neither Agree or Disagree
   - [ ] Agree
   - [ ] Somewhat Agree

6. I am confident in the system.
   - [ ] Strongly Disagree
   - [ ] Disagree
   - [ ] Somewhat Disagree
   - [ ] Neither Agree or Disagree
   - [ ] Agree
   - [ ] Somewhat Agree
<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>7. The system provides security.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Neither Agree or Disagree</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Somewhat Disagree</td>
<td></td>
<td>Strongly Agree</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Agree</td>
<td>Somewhat Agree</td>
</tr>
<tr>
<td>8. The system has integrity.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Neither Agree or Disagree</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Somewhat Disagree</td>
<td></td>
<td>Strongly Agree</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Agree</td>
<td>Somewhat Agree</td>
</tr>
<tr>
<td>9. The system is dependable.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Neither Agree or Disagree</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Somewhat Disagree</td>
<td></td>
<td>Strongly Agree</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Agree</td>
<td>Somewhat Agree</td>
</tr>
<tr>
<td>10. The system is reliable.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Neither Agree or Disagree</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Somewhat Disagree</td>
<td></td>
<td>Strongly Agree</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Agree</td>
<td>Somewhat Agree</td>
</tr>
<tr>
<td>11. I can trust the system.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Neither Agree or Disagree</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Somewhat Disagree</td>
<td></td>
<td>Strongly Agree</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Agree</td>
<td>Somewhat Agree</td>
</tr>
<tr>
<td>12. I am familiar with the system.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Neither Agree or Disagree</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Somewhat Disagree</td>
<td></td>
<td>Strongly Agree</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Agree</td>
<td>Somewhat Agree</td>
</tr>
</tbody>
</table>

*Please use the space below to provide additional information regarding the system.*