Master’s thesis presented to the Department of Psychology of the University of Basel for the degree of Master of Science in Psychology.

Lassoing the loop: An examination of factors influencing trust in automation

Zöe Fenn

Institute of Psychology

Department for Cognitive Psychology and Methodology

University of Basel

November 2020

Matriculation number: 15-064-520

Correspondence email: z.fenn@unibas.ch

Habilitated member of the Faculty of Psychology, examiner & supervisor: Prof. Dr. Klaus Opwis

Examiner: Dr. Florian Brühlmann

Submission date: 15.11.2020

Acknowledgments

This study was supported by the University of Basel from the department for Cognitive Psychology and Methodology. Thank you to Prof. Dr. Opwis for supervising this project. Many thanks to Daniel Gerber, Dr. Michael Fenn, Andrew Burgess, Prof. Dr. Rui Mata and Alexander Fenn for providing their support.

Declaration of scientific integrity

The author hereby declares that she/he has read and fully adhered to the Code for Good Practice in Research of the University of Basel.

Abstract

Multiple determinants influence operators’ decisions on whether to trust and rely on automated or semi-automated systems. The systems' reliability and transparency as well as the operator’s self-confidence are examples of such determinants. This study focuses on the effect of the level of automation (on the loop, in the loop), age (18 to 30, 45 to 55), and personality (Big Five) on operators' behavioural and self-reported trust. In a luggage screening task with 60 trials, 229 participants from the Amazon Mechanical Turk platform were asked to assess whether x-ray images contained a gun or a knife. Participants were assisted by a decision aid system. Depending on the experimental group, participants either had to actively decide on whether to follow the advice of the system or not, or they acted as supervisors and thus could only intervene to disagree with the system's decision. The main behavioural results suggest that participants on the loop supervising the system disagree more frequently than the participants in the loop, $\eta^2 = .026$, $p = .014$. Interestingly, these findings were not supported by participants' self-reported trust. No prominent age or personality effects were identified regarding participants' self-reported or behavioural trust. The level of automation appears to influence operators’ reliance on automation, but not self-reported trust. Several explanations, such as agency and boredom are discussed. Potential applications include considering how humans’ need for agency changes when designing interfaces between humans and automated systems with different levels of automation.

Automation and society

As amply illustrated by current events (home office and online retail expansion during the pandemic) technology and automation are playing an ever-increasing role in our lives. History has shown that the development of new technologies creates possibilities to improve the efficiency of certain forms of labour, e.g., horse and carriage to car. In recent years, systems that once needed manual control are evolving into semi-autonomous or completely autonomous systems. For example, factory workers are replaced by robots to perform repetitive tasks and to save money or autopilot functions take over the lead control of airplanes, enabling pilots to focus on other tasks (Naranji, Sarkani, & Mazzuchi, 2015; Wakefield, 2016).

A consequence of these developments is that the human has less control and less workload for the same task. However, as with most new inventions, potential dangers arise until the infrastructure is appropriately adapted (e.g., electricity plus fuses). On the one hand, there are many incidents in which human interaction with automated systems has gone terribly wrong and sometimes even ended with fatalities (Boudette, 2018; Tripti, 2018). On the other hand, there are examples where human intervention has saved the day. One of these incidents occurred in 1983, where Stanislav Petrov’s task was to monitor a satellite system that provided information on whether missiles were fired at the Soviet Union and to notify his superiors. On the 26th of September, he saw up to five missile launches on the satellite screen from the U.S. and had to make a quick decision on whether it was a false alarm. Luckily, he considered the situation a false alarm, since a first attack would almost certainly involve more than five missiles. The satellites had falsely identified the sun reflecting off cloud tops as flames from rocket engines. Petrov’s decision quite possibly prevented the start of World War III (Tegmark, 2017). In this incident, Petrov correctly trusted his rationale instead of the system's advice. Such incidents demonstrate the importance of understanding the underlying effects automated systems have on human behaviour, trust and cognition.

There are multiple suggestions on how to design beneficial interfaces (Norman, 2013). Most good interfaces are ergonomically designed for 95% of the people, yet for 5% they are not (TAHPI, 2015). Yet, if the failure of the design or system has more severe consequences (e.g., death or injury), it needs to be suitable for all users, and in some cases, users' trust must be properly calibrated for them to appropriately use certain systems.

Therefore, it is important to study individual differences in interaction with semi-autonomous and autonomous systems. Depending on the results, different measures may be appropriate for certain individuals, such as more training, prior information, or different interfaces. For instance, “smart” technologies for homes are becoming more and more popular. Some of these are purposely being developed to support older adults in their everyday life at home. Previous research suggests that older adults tend to over-rely on automation in comparison to younger adults (Schaefer, Chen, Szalma, & Hancock, 2016). Indeed, other characteristics may influence how older adults interact with these devices, such as risk preference or cognitive abilities. These factors should be investigated to understand whether older adults need to receive training on when to rely on and use these automized devices (Ezer, Fisk, & Rogers, 2008). In addition, not all tasks may be appropriate for all individuals and it may be worth considering whether to even design systems for older adults that require them knowing how to use some technologies, for example, an iPad.

With this frame of reference in mind, this thesis focuses on human decision making whilst interacting with semi-autonomous and autonomous systems. More specifically, it examines whether individual differences in trusting automation are present and to what extent different levels of automation and trust influence reliance.

Theoretical Background

This section introduces the reader to relevant constructs and draws attention to several research gaps in the domain of trust in automation. It is structured as follows: first, trust in automation is defined, followed by a theoretical background and measurement difficulties of this construct. Second, assumptions and associated factors that cause trust differences when interacting with automation are listed, providing the current state of research. Lastly, the rationale and overview of this study are given.

Trust in automation

«*Trust* can be defined as the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability» (Lee & See, 2004, p. 51).

This definition is interesting as it draws attention to two aspects of trust. First, trust requires an individual's attitude on whether an agent (e.g., computer) can help to achieve a goal or not to be positive. Second, it implies that if the individual trusts the agent the individual will most likely rely on the agent's advice or help. Furthermore, it is important to distinguish reliance from trust. In this context *reliance* is the act of depending on the system's decisions.

If trust and reliance are not adapted appropriately different issues arise: these can be classified as, disuse, misuse and abuse of automation (Parasuraman & Riley, 1997). Disuse of the system occurs when the individual has a high level of mistrust towards the system leading to low reliance on the system, whereas misuse results from an overreliance on the system due to decision biases. Automation abuse describes the decision to use automation without considering the consequences for humans, systems relation and performance. Malicious intent may be a reason for abuse.

Trust is only one important factor that influences participants' reliance on a system. Other factors include perceived workload, time pressure and perceived risk. Previous research has found difficulties in understanding to which degree each of these influences may shape the operator's reliance (Hergeth, Lorenz, Vilimek, & Krems, 2016). Measuring trust has also proved a challenge. On the one hand, questionnaires have been used to measure participants' trust in automation yet, like most measurement instruments for latent constructs, questionnaires have their limitations (e.g., social desirability bias; Hergeth et al., 2016; Razavi, 2001). On the other hand, some studies have attempted to capture trust from a behavioural perspective. Hergeth et al., for example, found a positive association between self-reported trust and operators' monitoring frequency of an automized system. Interestingly, multiple studies have observed a discrepancy between operators' verbal measures of trust and their behaviour, such as reliance (e.g., Rovira, McGarry, & Parasuraman, 2007). This discrepancy has also been found in other domains such as risk preference, for which Hertwig, Wulff, and Mata (2019) provide several possible explanations, for example, that behavioural measures in risk preference may capture transient states rather than stable traits. This may also be the case for individuals' trust in automation. A further explanation for this disparity may be a subconscious level of (mis)trust that does not appear in the questionnaires but in behavioural measures.

In addition, not all questionnaires may be measuring the same types of trust. In previous publications, trust in automation has been divided into three different types: dispositional trust, situational trust, and learned trust. These different types are described in Table 1. The interaction of these different types of trust produces the overall trust in an automatized system.

Operators' trust itself may vary to a large degree depending on the type that is looked at. Furthermore, operators' trust may additionally vary between different fields. Inagaki (2006) argued that a cooperative collaboration between humans and automation may be domain-dependent, indicating that factors such as operator's knowledge, training, or time

Table 1

Three different types of trust in automation.

Types of trust	Description
Dispositional trust*	Empirical findings support the first aspect of the above definition of trust, given that the self-reported predisposition of trust in automation varies from person to person (Lee & See, 2004). This attitude is assumed to be moderated by individual differences, such as age, personality, risk preference, task difficulty, self-confidence, need for control, agency and memory (Endsley, 2017; Schaefer et. al, 2016). Lee and See (2004) state that even a small difference in this predisposition may substantially affect the initial engagement with automation. For example, the results from Merritt and Illgen (2008) suggest that individuals with a high degree of dispositional trust are more prone to trust reliable systems, yet their trust may decline more heavily after system errors.
Situational trust	Research has shown that trust can intensify or diminish depending on external influences (e.g., automation reliability, level of automation, automation robustness, interface, transparency), and thus influencing how much the individual will depend on and therefore appropriately use the system (Chien, Lewis, Sycara, Kumru, & Liu, 2020; Kaur & Rampersad, 2018; Lee and See, 2004; Molnar et al., 2018; Parasuraman & Riley, 1997). Marsh and Dibben (2003), name this situational trust.
Learned trust	Apart from the dispositional trust and situational trust, Marsh and Dibben (2003) conceptually distinguish a further type of trust, which is the trust based on experience with a specific system. Learned trust will not vary in the moment, as it is based on past experiences with the system. Therefore, the operator's baseline of learned trust is different at the beginning of the interaction and change after the interaction is completed. Naturally, if an operator's learned trust is low, to begin with, there will be a stronger reluctance to rely on the systems support.

Note. Field marked with an asterisk was investigated in this study.

criticality may influence this collaboration. Indeed, it is conceivable that the severity of the consequences, which varies across domains, influences operators' collaboration, and situational trust. However, little research has compared operators' trust in automation across different domains (Pak, Rovira, McLaughlin, & Baldwin, 2017). In fact, Hoff and Bashir (2015) found that around 80% of 127 studies focusing on trust in automation were in the domains of military, security, or industry. These findings question the representativity and generalizability of scientific findings published. One explanation may be that other domains are too slow to adapt to new technology because of vested interest in paid jobs. Or, there is a reluctance to change the status quo due to threats of potential litigation. Pak et al. (2017) investigated this gap and gave participants 16 different scenarios entailing various automation reliabilities, domains (consumer electronics, banking, transportation and health), and stages in which the automated system provided support. Their results showed trust was highly dependent on the interaction of these three parameters and the user group, varying in age and civilian or military participants (students, cadets, and older adults).

In summary, the individual level of trust is based on several dispositional characteristics, prior experience, and situational factors, and these factors are presumably interdependent. Moreover, trust is one of many important components affecting individuals' automation reliance (Chien et al., 2020; Ho, Kiff, Plocher, & Haigh, 2005).

Factors influencing trust and reliance in interaction with automation

In the next section, research on the factors, level of automation, degree of reliability, age, and personality are introduced and their influence on trust and reliance are discussed.

Levels of automation. Depending on the level of automation, the operator has more or less active work. One extreme is a completely autonomous system, whereby the operator needs never to intervene. The opposite is complete manual control. The levels in-between depend on the amount of work and decision-making needed by the operator and system (Parasuraman, Sheridan, & Wickens, 2000).

Accordingly, levels of automation can be described as human “in the loop” (ITL), human “on the loop” (OTL), and human “out of the loop” (OOTL). OOTL describes an autonomous system, without the need for human labour. OTL stands for a system functioning autonomously, with an operator still overriding the system when needed. In this scenario, the operator has the role of a supervisor. The last term, ITL, describes a system in which the operator is an essential part of the process to accomplish a task. The operator must act at specific steps of the process and interacts with a semi-autonomous system (Scharre, 2018).

Different terminologies are used, describing specific phases in a decision-making process in which automation can replace an operator (Kaber & Endsley, 1997; Onnasch, Wickens, Li, & Manzey, 2014; Parasuraman et al., 2000); however, most share the same amount of stages and only differ in the naming. Parasuraman et al. (2000) named these phases: Information Acquisition, Information Analysis, Decision Selection, and Action Implementation. These four phases resemble the decision-making model developed by John Boyd (1987). Boyd’s Observe-Orient-Decide-Act (OODA) model was originally designed for the military domain, however, over time it has been adopted in further areas (Grant & Kooter, 2005).

Table 2

Decision stages and automation

Level of automation	Decision stages			
	Observe	Orient	Decide	Act
Manual control Each decision stage is explained here, and no automation assists in any stage (Grant & Kooter, 2005).	Operator interacts with the environment without any support and collects information.	Operator processes information interactively and includes it into a cognitive representation of the situation. Previous experiences, genetic heritage, new information and cultural background influence this stage.	The operator mentally decides and creates an internal hypothesis.	The operator acts on the decision and may receive feedback on the hypothesis.
ITL* At each stage automation may or may not assist the operator. The strength if assistance can vary.	The automation may assist by automizing the collection of data. Or cueing an unusual pattern in the data.	Automation may assist by processing and integrating the incoming data. For example, by calculating trends, or by providing suggestions.	The automation may assist by suggesting which decision to select.*	The automation may assist by deciding on when to act on the decision and execute it without human aid. *
OTL* At each stage there may or may not be the possibility to intervene. There must be at least one possibility.	An operator may intervene by stopping the data collection or rejecting certain information.	The system integrates the information by following a set of algorithms. An operator may intervene by, for example stopping this stage.	The system selects a decision, the operator may intervene by overriding the decision.*	The system acts on the decision. The operator may intervene by stopping the action.*
Fully autonomous	All actions are performed by the system. There is no possibility for the operator to intervene.			

In principle, a system can involve automation at all four phases, and for each phase, the amount of automized assistance can vary, this is what Parasuraman et al. (2000) refer to as the level of automation. Table 2 reflects the phases of the OODA in relation to the level of automation. Table 2 is not a summary of research in this domain but is intended to provide the reader with an overview of the theory discussed in this section and indicate the areas which are addressed in the study.

As Table 2 describes, the higher the level of automation the less active work is needed in one or more stages. A layperson might assume that a higher level of automation is consistently better. Past research, however, has shown that the higher the level of automation the higher the risk of negative effects on human performance (Endsley & Kiris, 1995). In this context, *performance* refers to how correct an operator carries out a task. Onnasch et al. (2014) conducted a meta-analysis with 18 studies on the mediating effect of levels of automation on performance. Their results suggest that performance benefits from a higher degree of automation if no failures occur; however if the support fails the performance declines with a higher level of automation. They found a similar but weaker pattern for workload. Not surprisingly, they additionally found the higher the level of automation the lower the situational awareness and the worse individuals performed in the event of system failures.

Research therefore suggests using automated systems that maintain the operator's engagement, either by keeping the human ITL or by developing a transparent interface, where the operator is informed on the situation (Endsley, 2017). One solution for this is *adaptive automation*. In this scenario, the automation engages the operator purposely by adding moments of manual control, for example, based on a set of time periods. More research is necessary to understand when such moments are needed (i.e., after how much time), which tasks should be selected for manual control to optimally stimulate engagement, and how to best architect them to maximise overall situational awareness. Additionally, care is needed in the design to avoid overwhelming the operator when switching between the two states. Of course, the operator must also trust this adaptive automation, which may be more demanding due to the varying states.

To summarize, a higher level of automation is more likely to reduce an operator's situational awareness. Depending on the reliability of the system, meaning how often the system is correct, it will be better or worse in identifying patterns or alarms correctly. So, as an operator, it is important to be able to identify false positives or false negatives. The story in the introduction with Stanislav Petrov underscores this point. However, when lacking situational awareness, which can arise as a result of boredom, mind wandering, distraction, personality differences in conscientiousness or others, such flaws may slip through, and depending on the task this could end with fatal results (Bailey, Scerbo, Freeman, Mikulka, & Scott, 2003; Gouraud, Delorme, & Berberian, 2017). Besides, when solely monitoring automation there is less cognitive engagement in the orient stage of the decision process and thus less understanding of the task and the surrounding context. This may lead to automation abuse or misuse, resulting in the operator's inability to intervene when needed. Overall, surprisingly little research has been published comparing the level of automation with trust (Lewis, Sycara, & Walker, 2018). This is somewhat remarkable since in military environments guidelines on how to interact with semi-autonomous systems and autonomous systems are provided (US Department of Defense Directive, 2012).

Degree of reliability. As mentioned above, the degree of reliability of a system can influence the trust of operators and thus their reliance and compliance with the system (McBride, Rogers, & Fisk, 2011). Parasuraman, Molloy, and Singh (1993) point out that in a highly reliable but not perfectly reliable decision aid system, the monitor's trust may not be well adjusted. As a result, if the system does not provide a warning in a critical situation, the operator may not realize the emergency due to overreliance. These failures of the system are so-called false negatives and may lead to misuse. Whereas false positives or false alarms will encourage under reliance and disuse. The latter failure is mostly considered more acceptable.

Rovira et al. (2007) investigated this assumed influence of reliability by using two different reliability levels to see how and whether this manipulation affected operators' performance (decision accuracy). This study is described in some detail to provide the reader with an example of how a study in this domain may be designed. Eighteen participants were invited to perform the “sensor to shooting” task, in which participants needed to identify the most dangerous enemy target. They were also required to choose a corresponding friendly unit to help in combat. The participants were able to select a decision aid system, that varied in the amount of decision automation reliability (60%, 80%) and the level of information automation (four different types). All participants went through all eight conditions. The researchers found several interesting findings. First, when the automated decision aid system was reliable, it improved operators' performance. This effect was mostly to be seen in the amount of decision time the operators needed for a task. Second, false automation advice trials led to worse performance under the 80% reliability condition for three different automation types relative to information automation (automation with no decision pointer). Third, false automation advice trials led to worse performance under the 60% reliability condition for all types of automation. This implies that when the decision aid system is under a certain threshold of reliability degree, it leads to poorer operators' performance, irrespective of the type of automation. Fourth, persons rely more on automation when it is more reliable, leading to greater compliance and therefore more detrimental effects when the automation fails. Fifth, the subjective trust ratings did not support the behavioural results, meaning there were no differences in the trust score for the two different automation reliabilities. In summary, their findings suggest that the reliability and the type of automation support may influence operators' automation usage, however, it may not influence operators' trust.

A meta-analysis on factors influencing operators’ trust contradicts the findings from Rovira et al. (2007). The results suggest that the reliability of the system matters and faults in the system affect operators' trust (Schaefer et al., 2016). One explanation of this is that operators adapt their trust to accommodate different levels of reliability, meaning a higher level of reliability would lead to a higher level of trust (Wiegmann, Rich, & Zhang, 2001). Although this effect is reduced when operators have prior knowledge of these faults (Lewandowsky, Mundy, & Tan, 2000). A potential reason for this is that this knowledge limits the uncertainty and consequent risk associated with automation usage (Lewis et al., 2018). Credibly perceived reliability may differ from actual reliability, as it could be biased by false expectations, expertise, or ignorance. This effect is also seen in the domain of algorithm decision-making (Burton, Stein, & Jensen, 2020). Previous research indicates that operators underestimate the true reliability of non-perfectly reliable automation, but that operators’ trust calibration improves after a considerable number of interactions with the automation (Wiegmann, 2002). Wiegmann et al. (2001) investigated, whether perceived reliability of a system with 80% reliability changes if the participant first interacts with the same system with a higher (100%) or lower reliability (80%). In both cases the participants rated the reliability of the system lower than the control group. The control group only interacted with an 80% reliable system. Their results suggest that trust in a decision aid system is more easily lost than regained.

Taken together, the higher reliability the better the operators’ performance. However, to date, the effect of the degree of reliability on trust is inconclusive, but findings suggest it is mediated by operators’ perceived reliability.

Age. Multiple studies suggest that older adults trust and rely on automation more than younger adults (Schaefer et al., 2016). This finding is supported from a behavioural perspective, as in the amount of agreement/disagreement with a system and from a subjective assessment, as in self-reported trust (Donmez, Boyle, Lee, & McGehee, 2006; Ho et al., 2005; McBride, Rogers, & Fisk, 2010). For example, McBride et. al (2011) examined to which degree the level of workload influenced individuals' compliance and reliance on a 70% reliable system, and whether this varied among older and younger adults. The researchers found that older adults complied and relied overall at a higher rate compared to young adults, while not changing this pattern when varying the workload. These results suggest that younger adults tend to trust their abilities more than the system. Conversely, that older adults do not trust their own capabilities and therefore rely more on the system than younger adults.

Ho et al. (2005) suggest that these age differences may arise due to cognitive changes, such as deficits in basic attention, memory, interpreting errors and learning. However, cognitive abilities may only be relevant to more complex choice environments (Hertwig et al., 2019). McBride et al. (2011) additionally mention that older adults may be less familiar with automation and therefore may be less aware of the concept of unreliability thus leading to deficits in interpreting errors. Indeed, Olson, O’Brien, and Rogers (2011) found that in general younger adults use a greater variety of technologies than older adults potentially leading to a better understanding of technologies due to more experience, but this pattern is domain-dependent.

A further explanation for an inadequate calibration of trust by older adults is their inability to form accurate mental models of the task environment. Olson, Fisk, and Rogers (2009) investigated this assumption and found that on average older adults do not form an accurate mental model. These individuals trust the automation more when the automation is wrong and make more misuse errors in comparison to individuals who have a moderate or highly accurate mental model. Still, others have found the opposite effect, that older participants tend to underutilize automized diagnostic aids (McCarley, Wiegmann, Wickens, & Kramer, 2003).

Additionally, several studies suggest that both younger and older adults tend to under rely on automation aid and show a bias for relying on themselves (Dzindolet, Pierce, Beck, & Dawe, 2002; Ezer et al., 2008). As illustrated, there are mixed findings on the age-related research and suggestions that the effects of age vary in distinct contexts (Hoff & Bashir, 2015).

Taken together, research suggests a tendency for a positive association between age and trust in automation, however, the topic is yet unresolved due to contradictory findings. Plausibly, age effects may vary across domains and across the systems being used and therefore resulting in inconsistencies between research findings.

Personality. Little research has been published on personality differences and decision-making regarding interaction with automation. However, the following section attempts to provide an overview of the research concerning personality differences and trust in automation.

Merritt and Ilgen (2008) looked at whether extraversion is positively related to the propensity to trust in automation, with the underlying assumption that extraverts may transfer their willingness to trust other humans to automated systems. They examined this with an x-ray screening task, in which 255 students visually searched for weapons in x-ray images of luggage. During the task half of the participants had the option to use and see a decision aid system and half always had the decision aid system in use. The researchers measured the participants’ pre- and post-task trust in the decision aid system. Their results suggest that extraverts have a greater propensity to trust machines than introverts do.

The study from Szalma and Taylor (2011) examined whether extraversion and the further Big Five personality traits influenced the amount of agreement with correct automated advice. The participants were invited to complete an uninhabited ground vehicle (UGV) task, in which their task was to decide whether they had seen a certain figure or not and if so they had to identify the figure from a list (e.g., terrorist, civilian). Depending on the study condition and block, they received a recommendation from a decision aid system, indicating which figure had been shown in the video clip. The researchers found a negative correlation between neuroticism and operators’ agreement with correct automated advice. The authors suggest that this could be because individuals with high levels of neuroticism are more cautious in their decision-making than others. Furthermore, in some experimental blocks, they found that the agreement of individuals with high neuroticism did not depend on the reliability of the automation. This indicates that the trait may reflect lower dispositional trust rather than situational or trust based on experience. This would support the assumptions from Hoff and Bashir (2015), that personality traits mostly influence dispositional trust. The researchers found no further associations between a personality trait and agreement.

Kraus, Scholz, Messner, Messner, and Baumann (2019) additionally found that higher levels of anxiety lead to distrust in automation. This supports the findings from Szalma and Taylor (2011), as anxiety is associated with neuroticism (Costa & McCrae, 1992). However, a very recent study found significant positive relations between dispositional trust and conscientiousness, agreeableness and openness. Moreover, the researchers found a negative relationship between extraversion and dispositional trust, these findings contradict the results from Merritt and Ilgen (2008) stated above. Interestingly, no association was found between neuroticism and dispositional trust (Ferronato & Bashir, 2020).

A further study examined the influence of Myers-Briggs personality types on automation usage. The results showed significant differences between judging and perceiving personalities in utilizing automated decision aids. More specifically, researchers found that perceivers had a higher tendency to have positive biases for using the automated decision aid system in comparison to judgers (McBride, Carter & Ntuen, 2012).

In the domain of algorithmic aids, some research suggests that confidence plays a considerable role in whether individuals are prepared to use algorithms (Burton et al., 2020). Dietvorst, Simmons, and Massey (2014) have found that purely seeing the algorithm making small mistakes systematically decreases the confidence in the model. Interestingly, however, watching a human conduct bigger mistakes does not consistently decrease the confidence in the human.

Lee and Moray (1994) found that self-confidence and trust influences operators' usage of automation. Their explanation is that if operators overestimate their capabilities they will tend to keep to manual control, whereas when operators' self-confidence is low and their trust is high, they will allocate more control to automation. Their findings also suggest that operators are predisposed to using manual control rather than complete automatic control and that over time operators are reluctant to alter their allocation strategy.

To conclude, little research examining the influence of personality types on trust in automation has been published and the findings of these are inconclusive. The interaction of trust and self-confidence is assumed to influence operators' allocation strategy when working with supervisory control systems.

Study overview and Rationale

The next section is structured as follows: First, the goal and the aim of the study is listed. Second, a short overview of the study is provided to introduce the hypotheses.

This study intends to respond to the frequent call for investigating the influence of individual characteristics on operators' trust in automation (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003; Lee & See, 2004). It focuses on age differences, due to the inconsistency among previous work; and on personality types, because of the limited amount of research performed on individual characteristics. This study also tests the influence of ITL and OTL on reliance and trust as little research has been published comparing OTL to ITL performance and trust (Hoff & Bashir, 2015). Most research has focused on altering the degree of reliability or has varied the automized support at different stages of the task, whereby keeping the operators ITL (e.g., Pop, Shrewsbury, & Durso, 2015).

This study investigated the influence of age and personality differences on the amount of disagreement with a system. It additionally investigated the influence of two different levels of automation on disagreement. Although disagreement with the system may not entirely reflect the individual's degree of trust in a system, it remains a reasonable measure to capture trust in automation from a behavioural perspective. Subjective trust ratings were also assessed to measure the individual estimation of the association.

Many studies have used rather complex paradigms involving several steps and information presentation between information collection and action implementation (Parasuraman, Bahri, Molloy, & Singh, 1991). This renders it difficult to assess the role of trust in automation on individual and age differences in decision-making. Thus, a simple paradigm was applied, similar to the one used in Merritt and Ilgen (2008).

More specifically, participants, who were either ITL or OTL, had to decide if weapons were hidden in x-ray images. The participants were supported by a system and asked to decide on whether or not to search the luggage. In contrast to Merritt and Ilgen (2008), the level of automation was manipulated instead of the degree of reliability.

The first hypothesis of this study is based on the majority of empirical findings, i.e., that older adults rely on and trust in automation more than younger adults:

1. Older individuals will disagree less with decision aid systems than younger individuals.

The second hypothesis below assumes that operators OTL will disagree less with the system due to less situational awareness:

2. Participants supervising an autonomous decision aid system (OTL) will disagree less than individuals interacting with a semi-autonomous decision aid system (ITL).

No specific hypotheses were made concerning the influence of personality differences on operators' trust in automation, given the sparse number of empirical findings on this topic. Consequently, the relations were solely investigated in an exploratory manner. Nevertheless, purely theoretical assumptions and empirical evidence from other domains could have been made, as in the study from Szalma and Taylor (2011) or Ferronato and Bashir (2020). For example, the trait conscientiousness is described by some of the following adjectives: efficient, dutiful, thorough, deliberate, and self-determined (John & Srivastava, 1999). High levels of conscientiousness are associated with a better job and academic performance (Barrick & Mount, 1991; Poropat, 2009). Therefore, it is imaginable that individuals with a high level of conscientiousness will perform better in supervising a system, as these are important attributes for accomplishing tasks.

Method

Pilot study

Before conducting the main online study, a pilot online study with 30 participants was conducted. As in the main study, participants were asked to decide with the help from a decision aid system, on whether to search or clear luggage (see procedure). Its main results showed that the decision task was too difficult, due to time limitations and difficulties in identifying weapons. For this reason, some images were replaced with easier images, and the time period per decision was extended from seven to nine seconds.

Participants

The online study was developed with the survey tool Unipark EFS survey (Questback GmbH, 2019). Individuals registered on the crowdsourcing platform, Amazon Mechanical Turk (Mturk), were recruited to participate in the study. Three hundred and twenty-six persons participated in the study, 92 participants were excluded, due to failure of attention checks or quality checks (see Appendix for more information) and five outliers were removed. Hence, 229 participants were included in the final analysis, 109 young adults (40 female, 67 male and two not further specified, range: 18–30 years) and 120 older adults (71 female, 48 male and one not further specified, age range: 45–55 years). On average participants needed 25 minutes to complete the study, and they received $2 USD for participating, with the opportunity to earn up to an additional 0.05 USD based on the data quality checks in the decision task. They could access the study by clicking on a link that directed them to Unipark.

To filter out bots and participants who are not attending to the purpose of the study, the following restrictions to participate in the study were set: a 95% approval rating from previous work (after a user completes a questionnaire or task the researchers administering the study need to approve or disqualify the work), the number of questionnaires approved greater than 100, aged between 18 to 30 or 45 to 55. These restrictions prevented the visibility of the study workers who did not satisfy the criteria. The recruitment was conducted in five batches and due to MTurk's age filtering settings, a batch was either for users aged between 18 to 30 or 45 to 55.

Procedure and Materials

The stimuli for the experiment were x-ray images of luggage, provided by Merritt, Shirase, and Foster (2020). The images contained different items, such as bottles, shoes, clothes, laptops and others. Dangerous items were restricted to guns and knives. Merritt et al.

(2020) assessed and reported the level of difficulty to identify a hidden knife or gun in each image. The difficulty grade for each image is the average probability to correctly judge whether a knife or gun is present. Each image was examined by 500 participants. For this study, 60 images were selected. Thirteen images had a correct decision rate between 80%−95%, eight between 60%−80%, 21 between 50%−60%, and 16 images between 30%−50%. The remaining two images had a correct decision rate of over 95% and were used as attention checks and therefore not considered in the analysis. Based on the results of the pilot study (*N* = 30) this difficulty level was deemed appropriate.

On the landing page of the online survey link, the participants were informed about the nature and duration of the study, data protection, and their payment. The participants could only proceed after giving their consent on the listed terms and conditions, they achieved this by clicking on an “I agree” box. Next, they were asked to report their appropriate age category, their gender, and their worker id (identification number on Mturk).

Before beginning with the main part of the study, the participants filled in a Big Five personality questionnaire with 44 items (John & Srivastava, 1999). Subsequently, the participants were randomly assigned to one of the two experimental groups (OTL or ITL). Depending on the group the participants read a different version of instructions, containing example images and the details of the task, noting that the computer's reliability rate was at 75% (this was also written in absolute terms). This information was provided to prevent a huge reduction of trust, given that prior knowledge on systems reliability has been found to reduce operators' mistrust (see the introduction for more information).

Next, they completed seven practice trials where they became familiar with the task (see Figure 1). They first saw the decision of the computer and after two seconds they saw the x-ray image. The participants had a total of seven seconds to decide, this number is based on past research and on the feedback from the pilot study (Szalma & Taylor, 2011). If the decision was correct, the participant would receive one point, otherwise they received zero points. Participants OTL could only actively disagree with the computer. More specifically, if the computer decided to search the luggage, then participants only had the possibility to click clear. If they however agreed with the computer, they were instructed not to click anything. After seven seconds the next image was presented. However, participants ITL had to actively click on an answer for each trial. If participants in the ITL group failed to click on time the image vanished after seven seconds, and the decision was recorded as a not available data entry. This may seem counterintuitive if one considers the concept of OTL and ITL (see Table 2). However, this design was selected to prevent the possibility of participants spending more time on each decision than the OTL group, and therefore potentially utilizing pattern search abilities rather than their intuitive decision on whether to trust the computer.

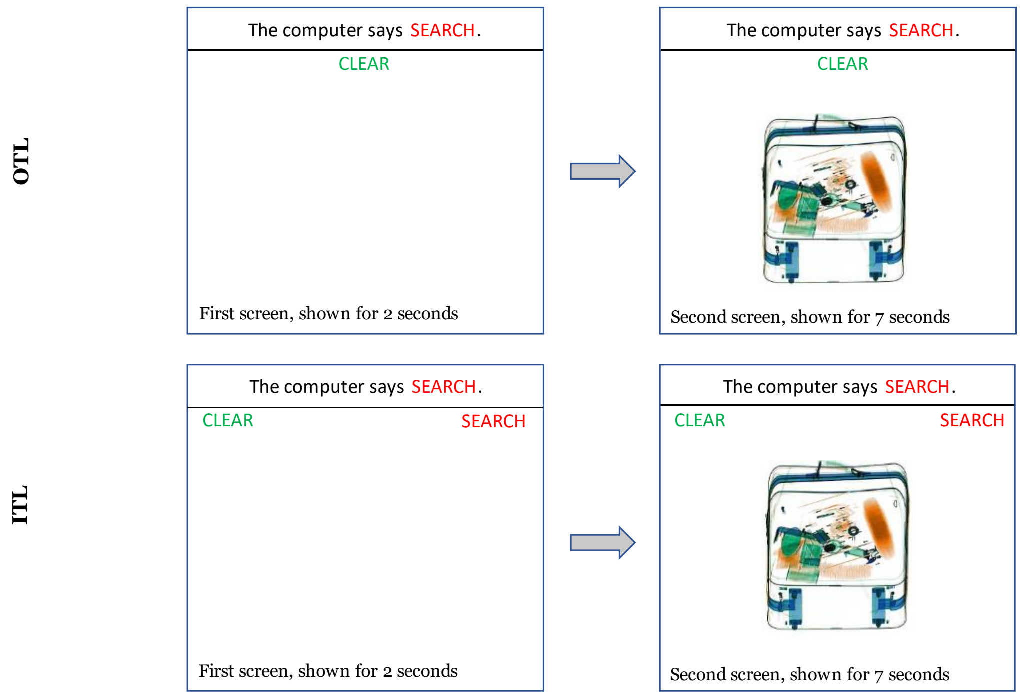

Figure 1. Schematic representation of the decision trials used in the study. During the first phase (duration: 2seconds) participants can only see the computer’s decision. In the second phase, an X-ray image additionally appears on the screen (duration: 7 seconds). The first row represents the trials from OTL and the second ITL. Only after pressing “CLEAR” or “SEARCH” did the participants ITL indicate their final decision. However, if they failed to decide it proceeded to the next trial, this was noted as a miss and the participants did not receive any points. Whereas in the OTL group if the contradictory button (“CLEAR” in this Figure) was not pressed then the computer’s decision was taken by default.

After completing 60 decision trials the participants were asked to fill in the *Trust in automated systems* questionnaire (Jian, Bisantz, & Drury, 2000), to assess the subjective trust in the system (i.e., the computer's decision). As a final step, they provided feedback on their internet stability, the number of missed trials, and their strategy. This was followed by an attention check, where they explained the task in a minimum of three sentences (see Appendix for more details). Only after completing the above steps did participants receive their achieved score and an individual code to prove their participation.

Variables

Level of automation. The study had two experimental groups to which the participants of the two age groups were randomly assigned. Therefore, the group variable had two levels namely "in the loop" (ITL, N = 131) and "on the loop" (OTL, N = 98). In the ITL condition, the computer merely suggested what the participant should decide, whereas in the OTL group the participants only needed to override the computer if they disagreed with the computer's decision.

Personality. The Big Five questionnaire by John and Srivastava (1999) was used to assess the participants' level of extraversion, neuroticism, agreeableness, conscientiousness, and openness. The questionnaire has 44 items that are measured on a five-point Likert scale ranging from strongly disagree (1) to strongly agree (5).

Age. Two age groups were tested in the study. The range for the younger adult group was 18 to 30 and in the older adult group, it was 45 to 55.

Trust. As mentioned above, two different assessments were performed. The participant's subjective trust was assessed, and the behavioural trust was assessed. The former of the two was measured with the questionnaire *Trust in automated systems* (Jian et al., 2000). This questionnaire has 12 items, which are measured on a seven-point Likert scale ranging from not at all (1) to extremely (7).

The behavioural measure of trust was used as the dependent variable, i.e., the extent of disagreement with the computer's choice. The *disagreement score* was calculated from the number of times the individual had disagreed with the computer’s recommendation, regardless of whether the participant's choice was correct. The maximum possible amount of disagreement was 58 and the correct amount of disagreement was 15. Furthermore, the dependent variable for the ITL condition was scaled to correct for an experimental design difference between the two groups. If during a trial the participants in the OTL group did not click on a button, then the decision of the computer was taken. However, if participants in the ITL group failed to click on a decision, this trial was missed (for more details see Appendix).

Score. This variable measured the performance of each participant. For each correct decision, the participant would receive one point and for each false decision, the participants would receive zero points. At the end of the study, the participants received their achieved score. Since two of the images were used as attention checks, the maximum score, that participants could achieve was 58 points. As mentioned above, five outliers were removed. An outlier was identified by multiplying the interquartile range by 1.5 and by adding this to the third or subtracting this from the first quartile. If an observation in the score variable was found outside of this newly defined range it was marked as an outlier.

Results

Descriptive statistics

To provide an overview of the measures, Table 3 shows the means, standard deviations, and the range of the variables measured during the study. Agreement and disagreement have the equivalent standard deviation since they are the inverse scores of one another. The maximum achieved score was 54.84 points, meaning that no participants achieved the maximum points possible (58).

Table 3

Mean, standard deviation, minimum and maximum of scaled variables measured during the study (N = 229).

Variable	Mean	SD	Min	Max
Disagreement	14.44	5.35	0	29.53
Agreement	43.56	5.35	28.47	58
Score	42.28	6.27	25	54.84
Big Five				
Extraversion	2.05	0.89	0	4
Neuroticism	1.53	0.96	0	3.88
Conscienciousness	3.04	0.74	1.2	4
Agreeableness	2.83	0.71	0.56	4
Openness	2.68	0.72	0.2	4
Trust in Automation	3.52	1.09	0.25	6

Analysis

To test the two hypotheses, i.e., to evaluate whether age or the experimental groups (OTL, ITL) determine the amount of disagreement with the computer's choice. A two-way Analysis of variance (ANOVA) with fixed effects was calculated. Before running the ANOVA test, the Levene-Test was conducted to test whether the data met the homogeneity assumption. No strong evidence for unequal population variances was found. Hence the variances were considered as equal, $F(3, 225) = 1.44$, $p = .23$.

Additionally, to determine the relationship between the participant's reported trust and the behavioural measure, a correlation was calculated. As in most psychological papers, an alpha value of .05 was used for all statistical tests, and the analysis was conducted in the statistics software *R* (R core team, 2014).

In the exploratory analysis personality traits were correlated with individuals' trust in automation. In addition, the results from the main analysis were explored in more detail.

The requirements for the interpretability of main effects were tested. These requirements depend on two criteria. First, the interaction should not be significant and, second, the rankings of the cell means should be the same, within each level of the other factor (Leonhart, 2009). The first assumption was tested with a type three ANOVA model. The results showed no significant interaction between age and group, $F(2, 225) = 2.25$, $p = .14$.

Additionally, when testing the second requirement, the order of the ranks for the factor age is different for each level of the group factor (see Figure 2). More specifically, older participants in the OTL group disagreed more than younger adults. Whereas ITL younger adults disagreed more than older adults. Therefore, only the main effect for group can be interpreted without limitations, since participants in the OTL disagreed more than the ITL group independent of their age. However, it must be noted that the graph does indicate a tendency for potential age differences (greater mean distance in the older group than the younger group).

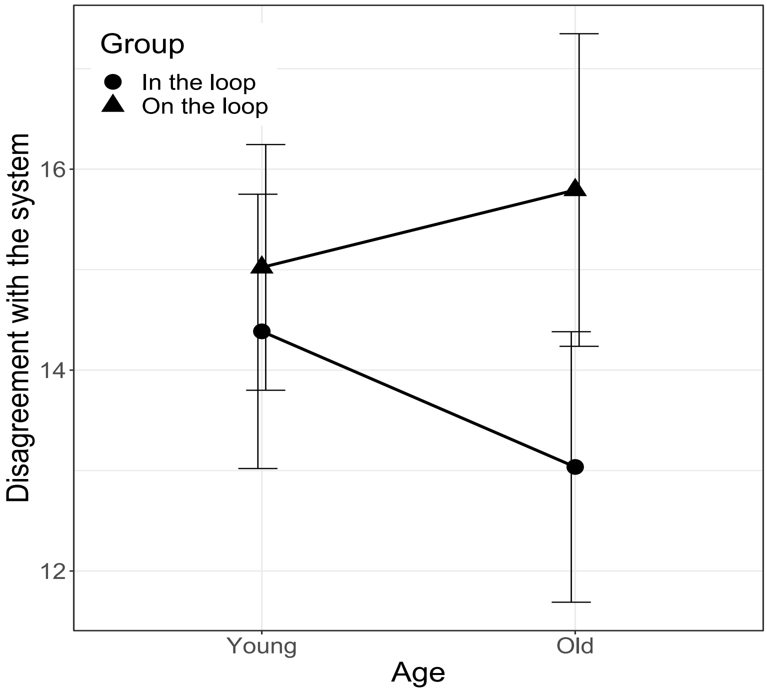

Figure 2. Interaction plot between age and group. The error bars represent 95% confidence intervals.

The results of the two-way ANOVA (type II) are represented in Table 4. The number of disagreements with the system was utilized as the dependent variable. Whereas the factors age and group functioned as independent variables, with the age levels young and old and with the group levels OTL and ITL. The results show a significant main effect of group on disagreement, $F(1, 226) = 6.16$, $p = .01$. In other words, the two different experimental groups differ in terms of the average amount of disagreement with the system. Indeed, the means of the two experimental groups revealed that disagreement with the system is significantly lower for the ITL group in comparison to the OTL group. However, the effect only explains 2.6% of the total variability in disagreement, which is accounted for by group differences.

There was a non-significant main effect of age on the amount of disagreement, $F(1, 226) = 0.40$, $p = .35$, meaning there was no difference between the two age groups regarding the amount of disagreement with the system. The interaction was not further examined since it does not answer the hypotheses. To investigate whether the number of counter-indications reflects the participants' subjective trust in the system, a correlation between the self-reported

Table 4

Descriptive statistics and the results from the two-way ANOVA showing the effect of age and group on disagreement.

	Mean	SD	N	SE	CI [lower]	CI [upper]	p	F	η^2	Power
Age							.35	0.40	-.002	.1
Young	14.65	4.93	109	0.471	13.71	15.58	-	-	-	-
Old	14.25	5.71	120	0.52	13.22	15.29	-	-	-	-
Group							.014*	6.16	.026	.7
OTL	15.44	4.97	98	0.50	14.44	16.43	-	-	-	-
ITL	13.70	5.51	131	0.48	12.74	14.65	-	-	-	-

$^{*}p < .05$

trust and the disagreement score was performed. Self-reported trust was weakly inversely related to the amount of disagreement with the system, $r = -.16$, $p < .01$, one-tailed.

Exploratory Analysis

In this section, the results from the main analysis are explored more closely and potential influences from different personality types are investigated.

Figure 3 shows the means of disagreement, agreement, and the overall score per group. The former two variables are divided into four sublevels: agree correct, agree false, disagree correct, and disagree false. As reported in the main analysis participants in the OTL group disagreed on average more than the ITL group (plot B). Additionally, plot C demonstrates that this tendency is present regardless of whether the decision was correct or wrong. However, when looking at agreement, the opposite trend is apparent. Participants in the OTL group (M = 42.6, SD = 4.97) agreed significantly less than participants in the ITL

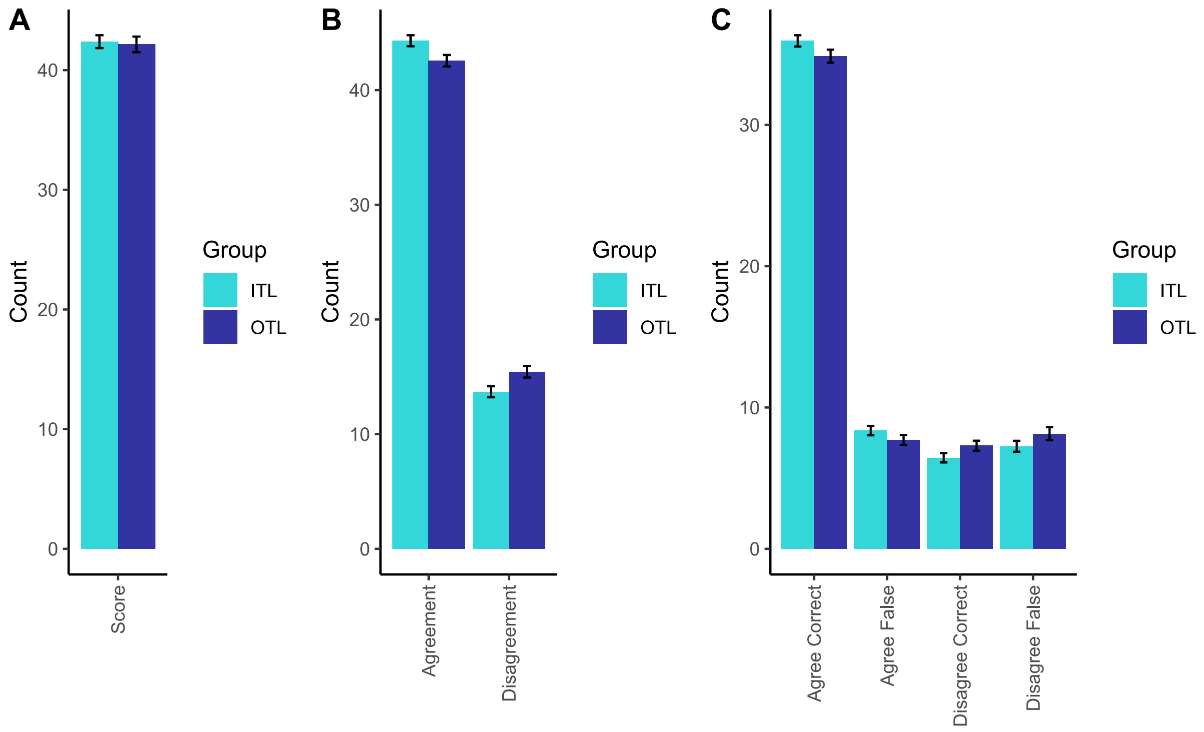

Figure 3. Figure three shows three different bar charts. Plot A shows the participants overall mean average score of the task, also divided by experimental group. Whereas plot B visualizes the mean amount of overall disagreement and overall agreement with the computer by group. Finally, plot C visualizes the breakdown of the average amount of disagreement and agreement by experimental group. The error bars represent standard errors. ITL = in the loop; OTL = on the loop.

group (M = 44.3, SD = 5.51) and therefore differed credibly between the two groups, t(219.12) = -2.5, p = .01. Consequently, both groups performed similarly on the task (plot A).

Owing to the sparse amount of literature on personality and trust in automation the influence of the Big Five factors was investigated in an exploratory manner. A multiple regression with five dimensions as independent variables and disagreement as the dependent variable did not show any significant findings. This was also the case when looking at the groups (OTL and ITL) independently. Additionally, each personality dimension was correlated with disagreement, correct agreement, and self-reported trust. There were no significant correlations with the disagreement score or for the correct agreement score. Yet interestingly, weak correlations were found with the self-reported trust. Figure 4 visualizes the relations between each personality dimension and self-reported trust with the system.

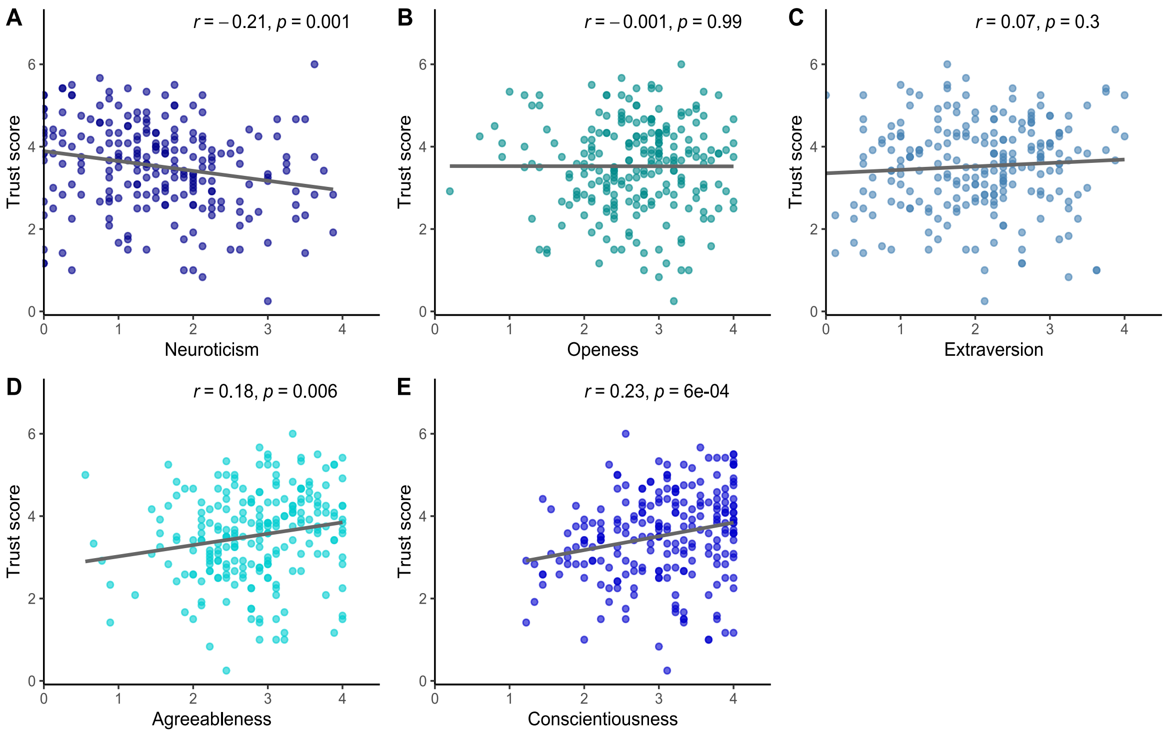

Figure 4. Pearson correlations between Big Five dimensions (score range: 0-4) and self-reported trust (score range 0-6).

Discussion

This study examined the influence of individual differences and the level of automation on trust in automation. More precisely, this paper investigated whether age or the level of automation (OTL and ITL) impacted the amount of disagreement with the system. The effect on trust by different personality types was examined in an exploratory manner. These influences were examined in an x-ray task. For the main analysis, a two-way ANOVA was applied.

Interpreting the main results

The results of the main analysis show no main effect for age, but a weak but significant effect for group (level of automation). The correlation between self-reported trust and disagreement was surprisingly low. The first hypothesis, that younger adults will disagree more than older adults can be rejected, as can the second hypothesis, that participants in the ITL group will on average disagree more than OTL. The following section will discuss the weak trends of the results.

The absence of an age effect contributes to the mixed findings among researchers, reported in the introduction. It supports the findings from Ho et al. (2005) suggesting that the pattern of automation usage is similar between younger and older adults. However, this study does not support the suggestion that younger and older adults under rely on automation aid, since the appropriate amount of disagreement, was at 15 and the mean disagreement for both age groups was 14.25 and 14.65 (Dzindolet et al., 2002; Ezer et al., 2008).

Nevertheless, the results are surprising, given that a simple choice environment was selected to allow age effects to come forward, implying that age per se may not be the leading force for differences. For example, a meta-analysis suggests that personality changes with age. Most changes occur during adolescence to young adulthood, however, some traits also vary in old adults, thus some effects may be counteractive. For instance, one of the findings from the meta-analysis was that trustworthiness is more prominent in old age, but openness to new experiences has been shown to reduce in old age (Roberts, Walton, & Viechtbauer, 2006; Sutter & Kocher, 2007). Therefore, these two elements may be working in two opposite directions and thus leading to no visible effect for age. Although no significant interaction effect was found, the interaction graph (Figure 2) does indicate different directions among the age groups across the two experimental groups. The influence of age on trust in automation appears to be a complex matter and should be further investigated.

A small effect was found between the two experimental groups. However, these findings are not aligned with previous research, which suggested that the higher the level of automation the less situational awareness and the more compliant the operator. Surprisingly, the results show an effect in the opposite direction. One could explain this effect with the concept of disuse implying a high mistrust among the participants in the OTL group and misuse among the ITL group (Parasuraman & Riley, 1997). However, there was no self-reported support for such a phenomenon and the average mean of disagreement by the OTL group (M = 15.4) was appropriate for the system's reliability (75%; 15 out of 58 trials were wrong). These findings are not aligned with previous research suggesting that operators under trust automation with imperfect reliability (Wiegmann, 2002).

A further explanation for this tendency may be boredom in the OTL group. In real-life settings, boredom may not even occur, given other stimuli present, for example, a phone call or a colleague asking how your weekend was. In this experimental setting, no such external factors should have influenced the participants. Although, naturally participants may have used their phone or watched TV whilst participating. Additionally, a recent review has found that the relationship between trust and automation decreases with the workload, given to the operator’s limited choice whether or not to rely on a system (Schaefer et al., 2016). Therefore, by increasing the informational and computational demands posed by the situation, group differences may emerge, although the influence of workload on trust in automation is still somewhat inconclusive.

Furthermore, qualitative feedback provided by the participants suggests that the screening task was too difficult. This is less likely to have resulted from the complexity of the task but from the difficulty in finding and identifying weapons. A supporting observation in favour of this feedback is that the OTL participants disagreed more independently on the verity of the system's advice. This might also explain the overall limited effects - i.e., participants may have relied on the decision aid system more than they would have normally done according to their dispositional trust. Thus, the situational trust may have dominated the dispositional trust.

Another interpretation for the group differences may be the operator's need for agency. Agency has been described as the experience of controlling one’s actions and thus external events (Gallagher, 2000). Research has shown that action selection is affected by the implicit sense of agency and that mere control influences behaviour independent of the outcome (Karsh & Eitam, 2015). Indeed, agency has been found to influence humans’ choice on whether to trust an algorithm’s assessment (Burton et al., 2020). Research also suggests that higher levels of automation go jointly with a decrease in agency (Berberian, 2019).

Considering this, participants on the loop may have disagreed more with the system solely because of their striving for agency and thus more control over the situation. This could be mitigated by designing an interface, where operators perceive to have more control over the situation. However, one should be careful not to reduce the transparency of the situation and task.

Interpreting the exploratory results

In the following section, the results from the exploratory analysis are discussed. Due to the nature of this analysis, the results should be interpreted with caution.

T-tests were conducted to compare differences between the two experimental groups across the different decision categories (see Figure 3). The results showed that the ITL group agreed more than the OTL group, independently of the correctness of the system's advice. Since the average misses in ITL were small (see Appendix) and the trial numbers were otherwise equal, this effect probably goes along with the inverse effect seen in disagreement.

No significant personality differences in disagreement or correct agreement were found. Thus, these results do not support the findings of disassociation between neuroticism and correct agreement. However, a weak negative trend for neuroticism and a weak positive trend for conscientiousness on self-reported trust were identified (cut off $r = -/+ 0.2$). The former tendency supports the association between anxiety and distrust in automation (Kraus et al., 2019). The latter tendency suggests that individuals with higher levels of conscientiousness found the system with a 75% reliability trustworthy. Conscientiousness is positively associated with greater task engagement, and potentially the participants' with higher levels of conscientiousness considered the decision aid system as reliable enough to perform well in the task (Matthews et al., 1999). Interestingly, these trends were not present when looking at the behavioural trust measure, disagreement. If conscientious individuals were to trust the system more, one would expect to observe this in the amount of disagreement or agreement with the system. However, conscientiousness is also positively associated with self-efficacy, which is associated with self-confidence, and as discussed in the introduction, individuals with higher levels of self-confidence tend to choose manual control over automation (Lee & Klein, 2002; McCormick, 2001). Therefore, the two components (trust and self-confidence) may be counteracting in individuals with higher levels of conscientiousness. Indeed, this may also be true for individuals with high levels of neuroticism (low level of trust and low level of self-confidence).

Yet, from a statistical perspective, the absence of the behavioural effect is not astonishing, given that the two scales (disagreement and self-reported trust) were only very weakly correlated, implying that the two measures may be capturing a different construct. Arguably, disagreement may reflect situational trust, whereas self-reported trust predominantly reflects dispositional trust. Since the participants completed the latter after the decision task the situational trust level may have affected the dispositional trust, and this could explain the reason for the weak effects. A further explanation may be due to the absence of significant negative consequences or rewards. Specifically, had a true or false decision of the participants' triggered a serious consequence in real life, personality, and age effects may have emerged.

Nevertheless, the absence of relation between the two scales is surprising, given that the questionnaire focuses on the automation the participants interacted with as opposed to their general trust in automation. Therefore, when examining individual differences future research should consider assessing the participant's self-reported trust before the interaction with a system. In addition, researchers should consider assessing participants' self-confidence.

Limitations

Several limitations directed towards the study are acknowledged:

1. The participants are individuals from the platform Mechanical Amazon Turk and they are therefore only a subset of the population and thus not representative. Even though the study entailed younger and older adults, the generalizability to the general public is quite likely to be limited. For example, younger people on Mturk may come, from a social perspective, from a better average demographic, since the willingness to work for a reward of 2 USD for older individuals presumably indicates a poor life outcome by the age of 45 to 55 but not necessarily for 18 to 30. Additionally, despite a bogus item in the

questionnaires, the data quality of the responses may be flawed, and this may explain the meagre effects.

2. The age differences between the two age groups (18 to 30 and 45 to 55) is lower than aimed. It was intended to include adults who are of working age, however, originally the ideal age groups would have been 18 to 30 and 55 to 65. Unfortunately, on account of Mturks possible age filters (45 to 55 or 55+), this was not possible. This may have contributed to the results, given that the age differences were too small for dissimilarities to emerge.
3. The comparability between the two experimental groups is somewhat reduced. Because participants in the ITL group could accidentally miss trials (by not selecting an answer) and participants in the OTL could not. The OTL, ITL construct in this study does not fully represent a real-life situation with these two systems. To deal with this issue the scores were scaled, nevertheless, the comparison should still be regarded with caution (see Appendix for more information).
4. The difficulty level of the decision task was not entirely appropriate. As mentioned above participants' feedback suggested that the task was too difficult. This is further supported by fact that the highest score out of 58 was 54.84 and the mean average score was 42.28 (see Table 3). This may have flawed the results, for example by suppressing the emergence of individual differences.
5. The participants received only a small reward for their participation and this was not contingent on the level of performance in the decision task. In the description of the study it was simply stated, that if participants did a good job they would receive a bonus, but the specific amount was not listed. This may have limited participants' motivation to perform well in the task, therefore leading to less engagement with the material.

Relevance and Suggestions for Future Research

The next paragraph aims to point out the relevance of further research for investigating individual differences on trust in automation, and the differences in decisions made by operators in an OTL or ITL situation. Additionally, recommendations are made on how to go about it.

As discussed, difficult tasks may lead to more agreement for humans ITL and more disagreement for operators on OTL. Potential explanations for this might be agency, boredom and mind wandering. Discovering whether this effect reveals itself in further settings is most important. This point can be illustrated by reference to the story in the introduction. The Soviet operator controlling the radar was (in this papers' terminology) in the loop and if his task to identify missiles was too difficult, he may have agreed with the system for the reason of not knowing better, with potentially disastrous consequences. Therefore, it is important to understand which decision environments support certain directions and tune the interfaces. Given that in some situations, disagreeing with a decision aid system could be safer than agreeing, further research should investigate whether characteristics such as gender or risk preference influence this behaviour. The results showed no evidence that age or certain personality traits disagree to a greater extent, yet these findings need to be stress-tested in other settings with more serious consequences, both in terms of risk or reward. Higher levels of risk and reward would be expected to enhance extrinsic motivation and also possibly intrinsic motivation, leading to greater engagement with the task potentially resulting in increased individual differences. Additionally, more research concerning the impact of the different levels of automation on trust is desired.

Naturally, the level of automation is not solely responsible for the intensity of trust an operator has in a system. Consequently, a deeper understanding of individual differences is required. On the one hand, how age, personality, and their interactions influence behavioural trust (reliance) and on the other hand their influence on dispositional trust. Mapping interdependencies between individual differences, such as age and personality, seems important when addressing the complex element of trust in automation. Moreover, it is necessary to further investigate these in more complex decision environments and across decision stages. Multiple factors such as transparency and understanding of the situation may interact differently at various timepoints with individual differences in trust.

The literature on trust in automation across domains is somewhat limited and more research is needed to understand individual willingness to interact with various systems in different contexts. Additionally, the interaction of the different subtypes of trust, self-confidence, and perceived reliability of the system should be examined with respect to their influence on reliance on the system across the different decision stages. Furthermore, how the different subtypes of trust vary over time after multiple interactions with a system should be investigated to a greater degree, for example, by using longitudinal studies.

In sum, future studies should systematically test the influence of boredom and agency on disagreement with semi-autonomous and autonomous systems. If these factors significantly influence feelings of responsibility, then this will impact behaviour and our interaction with autonomous systems, and the human-system touchpoints should be appropriately designed and tuned. In addition, the impact of the three types of trust, perceived reliability, and self-confidence should be investigated across decision stages.

Conclusion

In many domains, people are having to interact with more and more automized systems and, in some cases, they lack appropriate trust in the system. An experiment was conducted to investigate whether two different levels of automation influenced decision-making and whether older and younger adults arrived at different decisions. The main findings of the study were that age differences did not influence participants’ disagreement with the system. However, the level of automation did affect participants' disagreement with the system, participants on the loop disagreed on average more than participants in the loop. Of course, these results concern only one of many situations, where appropriate decision-making and trust in automation is needed. Considering this, future studies should test these results in further settings and investigate which influences create discrepancies in order to appropriately adjust certain technologies in the light of this knowledge.

References

Bailey, N. R., Scerbo, M. W., Freeman, F. G., Mikulka, P. J., & Scott, L. A. (2003). A brain-based adaptive automation system and situation awareness: The role of complacency potential. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *47*, 1048-1052. doi:10.1177/154193120304700901

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel Psychology*, *44*, 1–26. doi:10.1111/j.1744-6570.1991.tb00688.x

Berberian, B. (2019). Man-Machine teaming: A problem of Agency. *IFAC-PapersOnLine*, *51*(34), 118–123. doi:10.1016/j.ifacol.2019.01.049

Boudette, N. E. (2018, March 31). Fatal Tesla Crash Raises New Questions About Autopilot System. *The New York Times.* Retrieved from https://www.nytimes.com

Boyd, J. R. (1987). *A Discourse on Winning and Losing*. Retrieved from https://www.airuniversity.af.edu/Portals/10/AUPress/Books/B_0151_Boyd_Discourse_Winning_Losing.PDF

Burton, J. W., Stein, M., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, *33*, 220–239. doi:10.1002/bdm.2155

Chien, S.-Y., Lewis, M., Sycara, K., Kumru, A., & Liu, J.-S. (2020). Influence of Culture, Transparency, Trust, and Degree of Automation on Automation Use. *IEEE Transactions on Human-Machine Systems*, *50*, 205–214. doi:10.1109/THMS.2019.2931755

Costa, P. T. Jr, & McCrae, R. R. (1992*). NEO PI-R professional manual*. Odessa, FL: Psychological Assessment Resource.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144,* 114–126. doi:10.1037/xge0000033

Donmez, B., Boyle, L. N., Lee, J. D., & McGehee, D. V. (2006). Drivers' attitudes toward imperfect distraction mitigation strategies. *Transportation Research Part F: Traffic Psychology and Behaviour, 9*, 387–398. doi:10.1016/j.trf.2006.02.001

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies, 58,* 697–718. doi:10.1016/S1071-5819(03)00038-7

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2002). The Perceived Utility of Human and Automated Aids in a Visual Detection Task. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 44*, 79–94. doi:10.1518/0018720024494856

Endsley, M. R. (2017). From Here to Autonomy: Lessons Learned From Human–Automation Research. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *59*, 5–27. doi:10.1177/0018720816681350

Endsley, M., & Kiris, E. (1995). The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*, 381–394. doi:10.1518/001872095779064555

Ezer, N., Fisk, A. D., & Rogers, W. A. (2008). Age-Related Differences in Reliance Behavior Attributable to Costs Within a Human-Decision Aid System. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*, 853–863. doi:10.1518/001872008X375018

Ferronato, P., & Bashir, M. (2020). An Examination of Dispositional Trust in Human and Autonomous System Interactions. In M. Kurosu (Ed.), *Human-Computer Interaction.*

Human Values and Quality of Life (pp. 420–435). Cham: Springer International Publishing. doi:10.1007/978-3-030-49065-2_30

Gallagher, S. (2000). Philosophical Conceptions of the Self: Implications for Cognitive Science. *Trends in Cognitive Sciences*, *4*, 14–21. doi:10.1016/S1364-6613(99)01417-5

Gouraud, J., Delorme, A., & Berberian, B. (2017). Autopilot, Mind Wandering, and the Out of the Loop Performance Problem. *Frontiers in Neuroscience*, *11*, 541. doi:10.3389/fnins.2017.00541

Grant, T., & Kooter, B. (2005, June). *Comparing ooda & other models as operational view c2 architecture topic: C4isr/c2 architecture*. Paper presented at 10th International Command And Control Research And Technology Symposium on The Future Of C2, Virginia, USA. Retrieved from http://tinyurl.com/6bb5ne

Hergeth, S., Lorenz, L., Vilimek, R., & Krems, J. F. (2016). Keep Your Scanners Peeled: Gaze Behavior as a Measure of Automation Trust During Highly Automated Driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *58*, 509–519. doi:10.1177/0018720815625744

Hertwig, R., Wulff, D. U., & Mata, R. (2019). Three gaps and what they may mean for risk preference. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1766), 1–10. doi:10.1098/rstb.2018.0140

Ho, G., Kiff, L. M., Plocher, T., & Haigh, K. Z. (2005, November). *A Model of Trust and Reliance of Automation Technology for Older Users.* Paper presented at The AAAI-2005 Fall Symposium on Caring Machines: AI in Eldercare, Washington, DC. Retrieved from https://www.aaai.org/Papers/Symposia/Fall/2005/FS-05-02/FS05-02-008.pdf

Hoff, K. A., & Bashir, M. (2015). Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*, 407–434. doi:10.1177/0018720814547570

Inagaki, T. (2006). Design of human–machine interactions in light of domain-dependence of human-centered automation. *Cognition, Technology & Work*, *8*, 161–167. doi:10.1007/s10111-006-0034-z

Total Alliance Health Partners International. (2015). *Part C-Access, Mobility, OH & S*. In International Health Facility Guidelines. Retrieved from http://healthfacilityguidelines.com/GuidelineIndex/Index/Access-Mobility-OH-S

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

John, O., & Srivastava, S. (1999). The Big-Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives. In *Handbook of personality: Theory and research* (pp. 102–138). New York: Guilford Press. Retrieved from https://pages.uoregon.edu/sanjay/pubs/bigfive.pdf

Kaber, D. B., & Endsley, M. R. (1997). Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Safety Progress*, *16*, 126–131. doi:10.1002/prs.680160304

Karsh, N., & Eitam, B. (2015). I control therefore I do: Judgments of agency influence action selection. *Cognition*, *138*, 122–131. doi:10.1016/j.cognition.2015.02.002

Kaur, K., & Rampersad, G. (2018). Trust in driverless cars: Investigating key factors influencing the adoption of driverless cars. *Journal of Engineering and Technology Management*, *48*, 87–96. doi:10.1016/j.jengtecman.2018.04.006

Kraus, J., Scholz, D., Messner, E.-M., Messner, M., & Baumann, M. (2019). Scared to Trust? Predicting Trust in Highly Automated Driving by Depressiveness, Negative Self-Evaluations and State Anxiety. *Frontiers in Psychology*, *10*, 2917. doi:10.3389/fpsyg.2019.02917

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*, *40*, 153–184. doi:10.1006/ijhc.1994.1007

Lee, S., & Klein, H. J. (2002). Relationships between conscientiousness, self-efficacy, self-deception, and learning over time. *Journal of Applied Psychology*, *87*, 1175–1182. doi:10.1037/0021-9010.87.6.1175

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*, 50–80. doi:10.1518/hfes.46.1.50_30392

Leonhart, R. (2009). *Lehrbuch statistik: Einstieg und vertiefung.* Bern: Hans Huber.

Lewandowsky, S., Mundy, M., & Tan, G. (2000). The dynamics of trust: Comparing humans to automation. *Journal of Experimental Psychology. Applied*, *6*, 104–123. doi:10.1037/1076-898X.6.2.104

Lewis, M., Sycara, K., & Walker, P. (2018). The Role of Trust in Human-Robot Interaction. In H. A. Abbass, J. Scholz, & D. J. Reid (Eds.), *Foundations of Trusted Autonomy* (pp. 135–159). Cham: Springer International Publishing. doi:10.1007/978-3-319-64816-3_8

Marsh, S., & Dibben, M. R. (2003). The role of trust in information science and technology. *Annual Review of Information Science and Technology*, *37*(1), 465–498. doi:10.1002/aris.1440370111

Matthews, G., Joyner, L., Gilliland, K., Campbell, S. E., Falconer, S., & Huggins, J. (1999). Validation of a comprehensive stress state questionnaire: Towards a state "Big Three?" In I. Mervielde, I.J.Dreary, F. DeFruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 335–350). Tilburg, the Netherlands: Tilburg University Press.

McBride, M., Carter, L., & Ntuen, C. (2012). The impact of personality on nurses' bias towards automated decision aid acceptance. *International Journal of Information Systems and Change Management*, *6*, 132–146. doi:10.1504/IJISCM.2012.051148

McBride, S. E., Rogers, W. A., & Fisk, A. D. (2010). Do Younger and Older Adults Differentially Depend on an Automated System? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *54*, 175–179. doi:10.1177/154193121005400208

McBride, S. E., Rogers, W. A., & Fisk, A. D. (2011). Understanding the Effect of Workload on Automation Use for Younger and Older Adults. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *53*, 672–686. doi:10.1177/0018720811421909

McCarley, J. S., Wiegmann, D. A., Wickens, C. D., & Kramer, A. F. (2003). Effects of Age on Utilization and Perceived Reliability of an Automated Decision-Making aid for Luggage Screening. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *47*, 340–343. doi:10.1177/154193120304700319

McCormick, M. J. (2001). Self-Efficacy and Leadership Effectiveness: Applying Social Cognitive Theory to Leadership. *Journal of Leadership Studies*, *8*(1), 22–33. doi:10.1177/107179190100800102

Merritt, S. M., & Ilgen, D. R. (2008). Not All Trust Is Created Equal: Dispositional and History-Based Trust in Human-Automation Interactions. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*, 194–210. doi:10.1518/001872008X288574

Merritt, S., Shirase, L. and Foster, G. (2020). Normed Images for X-ray Screening Vigilance Tasks. *Journal of Open Psychology Data*, 8(1), 2. doi:10.5334/jopd.48

Molnar, L. J., Ryan, L. H., Pradhan, A. K., Eby, D. W., St. Louis, R. M., & Zakrajsek, J. S. (2018). Understanding trust and acceptance of automated vehicles: An exploratory simulator study of transfer of control between automated and manual driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *58*, 319–328. doi:10.1016/j.trf.2018.06.004

Naranji, E., Sarkani, S., & Mazzuchi, T. (2015). Reducing Human/Pilot Errors in Aviation Using Augmented Cognition and Automation Systems in Aircraft Cockpit. *AIS Transactions on Human-Computer Interaction*, *7*, 71–96. doi: 10.17705/1thci.00066

Norman, D. A. (2013). *The design of everyday things* (Rev. ed.). New York: Basic Books.

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

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

Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human Performance Consequences of Stages and Levels of Automation: An Integrated Meta-Analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *56*, 476–488. doi:10.1177/0018720813501549

Pak, R., Rovira, E., McLaughlin, A. C., & Baldwin, N. (2017). Does the domain of technology impact user trust? Investigating trust in automation across different consumer-oriented domains in young adults, military, and older adults. *Theoretical Issues in Ergonomics Science*, *18*, 199–220. doi:10.1080/1463922X.2016.1175523

Parasuraman, R., Cosenzo, K. A., & Visser, E. D. (2009). Adaptive Automation for Human Supervision of Multiple Uninhabited Vehicles: Effects on Change Detection, Situation Awareness, and Mental Workload. *Military Psychology, 21*, 270–297. doi:10.1080/08995600902768800

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *30*, 286–297. doi:10.1109/3468.844354

Parasuraman, R., Molloy, R., & Singh, I. (1993). Performance Consequences of Automation Induced Complacency. *International Journal of Aviation Psychology, 3*, 1-23. doi:10.1207/s15327108ijap0301_1

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*, 230–253. doi:10.1518/001872097778543886

Pop, V. L., Shrewsbury, A., & Durso, F. T. (2015). Individual Differences in the Calibration of Trust in Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*, 545–556. doi:10.1177/0018720814564422

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, *135*, 322–338. doi:10.1037/a0014996

Questback GmbH. (2019). EFS Survey (Version EFS Herbst 2019) [Online software]. Retrieved from https://www.unipark.com/en/

R Core Team. (2020). R: A Language and Environment for Statistical Computing (Version 4.0.0) [Computer software]. Vienna, Austria.

Razavi, T. (2001). Self-report measures: An overview of concerns and limitations of questionnaire use in occupational stress research [Discussion Papers in Accounting and

Management Sciences]. Southampton: University of Southampton School of Management.

Roberts, B., Walton, K., & Viechtbauer, W. (2006). Patterns of Mean-Level Change in Personality Traits Across the Life Course: A Meta-Analysis of Longitudinal Studies. *Psychological Bulletin*, *132,* 1–25. doi:10.1037/0033-2909.132.1.1

Rovira, E., McGarry, K., & Parasuraman, R. (2007). Effects of Imperfect Automation on Decision Making in a Simulated Command and Control Task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*, 76–87. doi:10.1518/001872007779598082

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 58*, 377–400. doi:10.1177/0018720816634228

Scharre, P. (2018). Machines that kill: What is an autonomous weapon? In *Army of None: Autonomous Weapons and the Future of War* (pp. 35–56). New York: W.W.Norton & Company.

Sutter, M., & Kocher, M. G. (2007). Trust and trustworthiness across different age groups. *Games and Economic Behavior, 59*, 364–382. doi:10.1016/j.geb.2006.07.006

Szalma, J. L., & Taylor, G. S. (2011). Individual differences in response to automation: The five factor model of personality. *Journal of Experimental Psychology: Applied*, *17*, 71–96. doi:10.1037/a0024170

Tegmark, M. (2017). *Life 3.0.: Being Human in the Age of Artificial Intelligence*. Great Britian: Peguin Books.

Tripti, L. (2018, November 28). Lion Air's deadly flight was a 13-minute struggle between man and machine. *Quartz*. Retrieved from: https://qz.com/1477182/lion-air-crash-report-details-pilot-battle-with-boeing-737-max/

US Department of Defense Directive. (2012). *Autonomy in Weapon Systems* (No. 3000.09).Retrieved from https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf

Wakefield, J. (2016, May 25). Foxconn Replaces '60,000 Factory Workers with Robots'. *BBC News.* Retrieved from https://www.bbc.com

Wiegmann, D. A. (2002). Agreeing with Automated Diagnostic Aids: A Study of Users' Concurrence Strategies. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *44*, 44–50. doi:10.1518/0018720024494847

Wiegmann, D. A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. *Theoretical Issues in Ergonomics Science, 2*, 352–367. doi:10.1080/14639220110110306

Appendix

Attention checks and quality checks

The online study contained several different attention checks to filter out bots and participants who simply clicked through the study without reading the instructions. Since the more studies participants complete, the more money they can receive, this behaviour is relatively frequent among Mturkers. Therefore, three different types of attention checks were made part of the experimental setup. First, a bogus item was placed in the first questionnaire (Big Five). The item stated: “This is an attention check. Please click on strongly agree.” Second, during the decision task, participants were confronted with two very easy images for which the recommendation of the computer was wrong. Participants, therefore, had to disagree with the system's advice. Last, at the end of the study, participants were asked to fill in two questions. One question asked the participants to describe the task. The second question tested whether the participants understood the task. Under each question, the nature of the question was clearly stated (attention check). Below are the two questions, as in the study:

1. “What did the task you had to perform look like? Please describe this in a minimum of 3 sentences. Your description should include the objects that you had to search for and the amount of time you had per decision.”
2. “During the trials, what happened if you did not click on an option (search or clear)?”

In the description of the study, it was clearly stated that the study had attention checks and that these will influence whether their work will be approved, 92 participants were excluded due to the attention checks. It was also stated that the participants should be able to understand English. If the bogus item was false, the work of the participant was not approved because it was not possible to complete the questionnaire without seeing the bogus item

(31/92). If the participants failed on both easy images, their overall response pattern was checked, and if it seemed unusual (e.g., only agreeing) the participants' work was not approved (13/92). The written responses at the end of the study were read and if it was clear from the description that the participants had not read the instructions then their work was not approved (34/92). For example, for the first question, one participant wrote “very interesting” and another copied and pasted the question into the response field. Also, if participants used a smartphone or tablet to complete the study, their work was not approved (9/92). (This was also stated in the description of the study and the type of device was recorded by the survey provider, Unipark). In addition, if the participants did not provide the correct participation code that was given at the end of the study they were excluded (2/92). Three participants clicked on two different age categories which contradicted each other and were therefore excluded (e.g., 45 to 55 and 55+).

Missed trials

Participants OTL could only actively disagree with the computer. More specifically, if the computer decided to search the luggage, then participants could only click clear. If they however agreed with the computer, they were instructed not to click anything. After seven seconds the next image was presented. However, participants ITL had to actively click on an answer for each trial. Although, if participants in the ITL group failed to click on time the image vanished after seven seconds, and the decision was recorded as not available data entry and was, therefore, a missed trial. The management of this difference between the two groups is explained below.

Out of 131 participants, 43 participants completed all decision trials. As Figure 5 demonstrates, a fair number of participants missed at least one decision trial (*n* = 88).

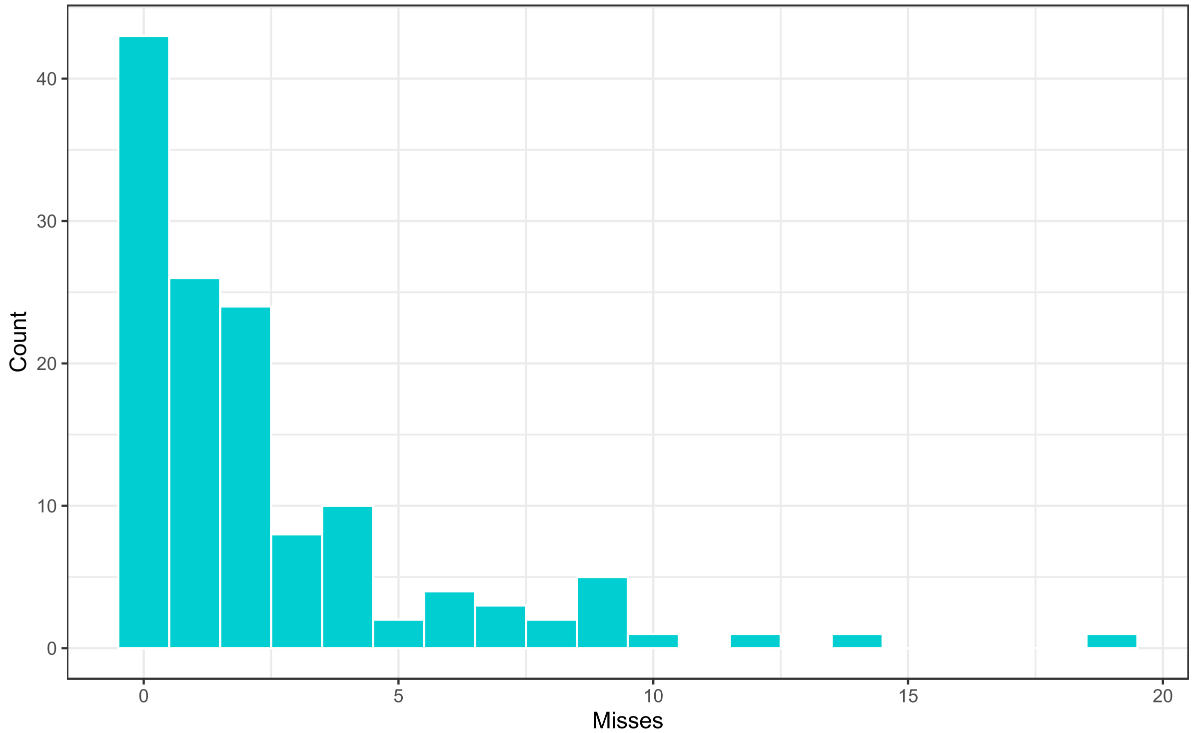

Figure 5 This figure shows the amount of times participants ITL missed a decision trial during the whole study.

This bias was corrected as follows.

$$Measure \cdot \frac{Total}{Total - Misses}$$

The measures disagreement, agreement, and score (as in performance) were scaled with where *Total* is the maximum amount of points that participants could achieve (58 points). *Misses* is the number of trials missed. *Measure* is the amount of disagreement respectively agreement or score, achieved for a variable.