Algorithmic Authority

Motives behind relinquishing decision-making to artificially intelligent systems

Evelien Mols
Graduation Thesis, 31 January 2020
Media Technology MSc program, Leiden University
Supervisors: Peter van der Putten, Francien Dechesne
e.mols@umail.leidenuniv.nl

***Abstract* — The application of Artificial Intelligence is increasing, but research about the predictors for acceptance of such technology is lacking. More specifically, reasons for granting artificially intelligent systems authority remain speculative. This research investigates what factors influence granting authority to algorithmic aid during a decision task. First, a conceptual framework was built to provide insights into the elements contributing to algorithmic authority. Next, an experiment was developed and conducted where one had to decide on the dismissal of employees. During these decisions participants had to accept or reject advice offered by an algorithmic aid. Four experimental groups received different background information about the algorithm. We analyzed the responses of 212 respondents. Results show that known historical use of the algorithm, self-confidence of the aid or background information regarding the qualification of the developers do not have a direct effect on algorithmic authority in comparison with the control condition. Information about social usage of the algorithm has, contrary to the hypothesis, a significant negative effect on the amount of granted authority to the algorithm. The content of the background information may alter the perceived task complexity and quality of the algorithm. A negative correlation was found between perceived task complexity and authority granted to an algorithm. Results further indicate a positive correlation between the perceived quality of an algorithm and the amount of granted authority. Age seems to have a negative effect on algorithmic reliance. This research contributes to our knowledge of human interaction with algorithm recommendations during dismiss decisions.**

***Index Terms*— Algorithmic Authority, Artificial Intelligence, Decision-making, Human-Computer Interaction.**

I. Introduction

In 1949 George Orwell wrote *1984*: a book about a dystopian future where human behavior is preprogrammed. In 2020, automated systems manage which news articles we read, which roads we take, which series to watch on Netflix and which person to date. If we put all our trust in the hands of technology, who is exactly controlling who? In other words; how accurate was Orwell with his dystopian future?

Artificially Intelligent (AI) systems are aiding us during many different tasks. We live in the era of *big data*, where algorithms increasingly affect our daily life. How much control do we give to these algorithms through automated decision-making; are we completely obedient? In this research we will investigate *algorithmic authority*. Algorithmic authority refers to the power of algorithms to manage human action and influence the accessibility of information (Lustig et al., 2016).

Algorithms play an increasingly important role in the working environment. Companies base more and more important decisions on the outcomes of big data, such as which people to hire and where to target their advertising (Hale, 2018). Makridakis (2017) talks about the forthcoming *AI revolution* and predicts both great advantages and challenges for all aspects of our society with the rise of AI technology. The key values we need to bear in mind in this era of algorithmic advancement, are to prioritize human control, dignity, fairness and accuracy (Araujo et al., 2019).

Risks regarding AI implementations are portrayed in recent news articles and involve issues concerned with systematic biases and lack of explainability and transparency (Dastin, 2018) (Weijer, 2018). Although we are increasingly surrounded by algorithmic agents that provide us with personalized content, research into the normative implications of such algorithmic authorities is lacking (Bodo et al., 2017). It is clear that we give our trust to seemingly intelligent systems and even base our lives on algorithms, as is the case with for instance autopilot in aviation (Logg, Minson & Moore, 2019). With the rising power of big technology companies that develop algorithms based on their access to massive amounts of data, it is important to find the motives behind granting intelligent systems authority. Why do we trust them and to what extent do we let them persuade us? If we do not understand how people exactly use algorithmic information during decision-making, both organizations and individuals risk missing possible opportunities that technological advances offer us (Logg et al., 2019). Liu, Helftenstein and Wahlstedt (2008) even vow for using technology for more than improving human efficiency, and state that it is of growing importance that technological solutions should be able to influence human desires. Opposing public concerns towards automated decision-making are manipulation, risk and unacceptable outcomes (Araujo et al., 2019).

If we can find out what factors play a role in giving authority to technology, this could have major implications for both social and ethical aspects concerning the use of AI. We may use this knowledge to define how artificially intelligent systems should further be developed and designed for interaction. The question we address during this research is *What factors influence granting authority to algorithmic aid during a decision task?* To answer this question we will first explore

existing literature on granting authority to systems. With the found literature, we will build a conceptual framework to highlight the possible determinants for granting authority to algorithms. Next, we will investigate to what extent each of those factors play a role during a decision task, where subjects can either accept or reject the advice of an algorithm. The remaining part of the paper will proceed with the analysis of the experimental outcomes, a discussion on those results and the conclusion, where we present the answer to the research question.

II. Background and related work

In this part we will examine the existing literature on algorithms and authority. We will take a holistic approach to uncover all factors that may contribute granting authority to decision aid within human-agent interaction. Throughout this thesis, the term agent will be used to refer to an artificial agent, such as a robot, computer or algorithm. It is necessary to clarify that within this research granting authority is defined as adopting to, and thus accepting, a recommendation of an agent. We will further elaborate on this decision in sub-section *C. Authority*. The goal of this literature research is to discover contributing factors for using technology in general, but also to provide a more in depth analysis on what is needed to trust agents and grant authority to them during a decision task.

A. Acceptance and use of technology

A considerable amount of literature has been published on factors influencing the use behavior of a certain technology. This resulted in a commonly used framework called the '*unified theory of acceptance and use of technology'* (UTAUT) (Venkatesh et al., 2003). In this framework there are two dependent factors, namely behavioral intentions and use behavior. Four independent variables influencing usage behavior have been identified: effort expectancy, performance expectancy, social influence and facilitating conditions. Gender, age, experience and voluntariness have found to moderate these variables. The UTAUT model is displayed in Figure 1.

B. Algorithms

An algorithm is a process to be followed in problem-solving operations or calculations. Within Machine Learning (ML), the most dominant and known technique of AI, algorithms are being created that can learn from data. A ML- algorithm is able to independently find patterns in data to build models and give predictions. The application of ML has substantially increased over time, due to the rise in available data (van Belkom, 2019). Literature has emerged that offers contradictory findings regarding the use of algorithms, which we will elaborate on further in this section. We will also discuss findings on social influence on the level of algorithm adoption and briefly address what is meant with algorithms' black box.

Algorithm aversion

During prediction tasks, such as forecasting a students' success by an admission office, evidence-based algorithms can be useful to factor a model to predict success. In many of such tasks, algorithms are more accurate than human forecasters (Silver, 2012). However, people tend to show a sense of *algorithm aversion*; the preference for human forecasts over those of algorithms (Dietvorst, Simmons & Massey, 2014). In a later research, Dietvorst et al. (2016) found a way to reduce this aversion by giving people some control over the forecasting algorithm, by giving them the freedom to slightly modify the outcome. This resulted in a greater belief that the algorithm was superior, and consequently participants were more likely to use it. Dietvorst et al. (2016) told the participants of their experiment that the model was based on data of hundreds of past students and was sophisticated, put together by thoughtful analysts. They revealed that people are unlikely to use the forecast of an algorithm, after receiving feedback that it is imperfect. Seeing an algorithm err has a much higher negative effect on the level of confidence in a system, than the level of confidence in other humans is affected after they make a mistake (Dietvorst et al., 2014).

Although the effect of immediate feedback is important for the theoretical framework of algorithmic advisors, there are many real-life cases where such feedback is not available. For instance during forecasts about political events or climate

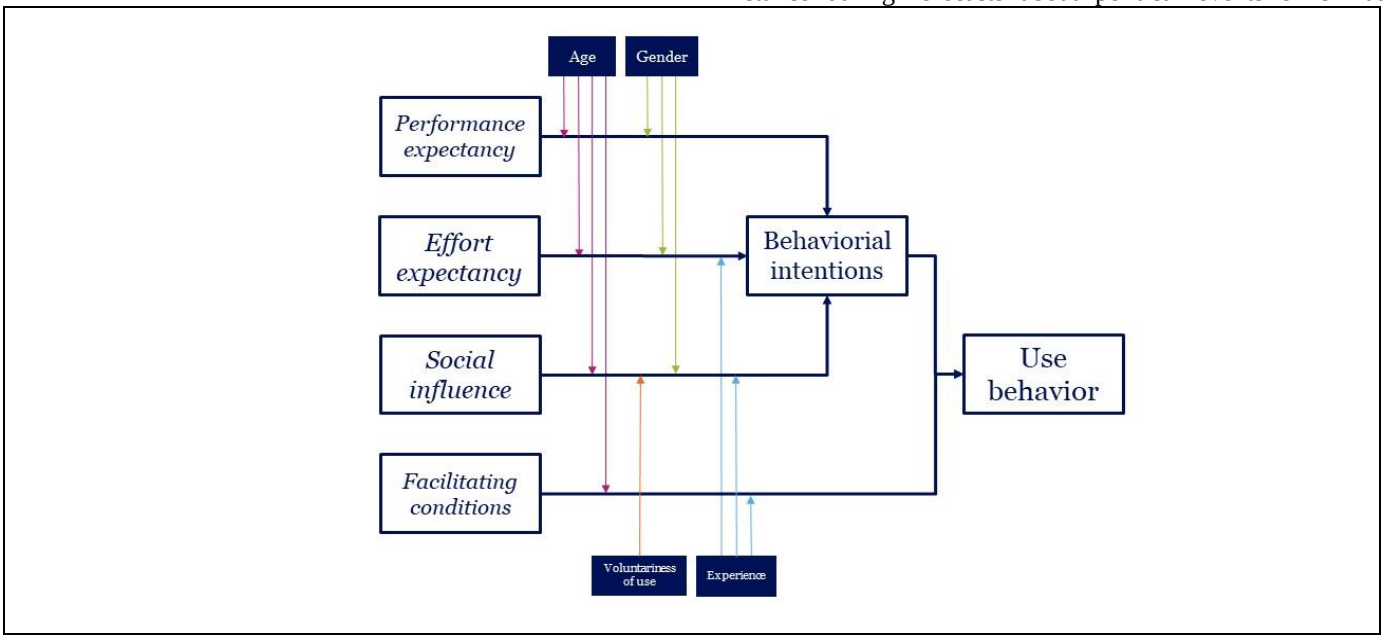

Figure 1. UTAUT framework (Venkatesh et al., 2003).

change (Logg et al., 2019). Dietvorst et al. (2016) also found that people are less likely to use an algorithm during an all-or-nothing setting. This was similarly found in another research, where Emmen (2015) showed that participants less often chose to use computer assistance during a chess game when the decision was irreversible. This might be explained by the desire for humans to keep a certain level of autonomy (Agrawel & Williams, 2017).

Algorithm appreciation

Counterintuitive to the effect of algorithmic aversion, Logg et al. (2019) found that under specific conditions people prefer algorithmic advice over human advice. They called this effect *algorithmic appreciation*. During one of the tasks within the experiment of Logg et al. (2019) people's reliance on algorithmic judgment was compared to their self-generated judgments; subjects had to choose between their own judgment versus the judgment of an algorithm, without information about prior performance. Subjects had to estimate ranks and chose algorithm's estimates over their own estimations for 66% of the time. Interestingly, they found that professionals rely less on advice from algorithms than their lay counterparts. We find the same result in the research of Emmen (2015) where people with a higher proficiency in the specific task were less inclined to use computer aid. This effect is also portrayed in the UTAUT model, where we can see the moderating effect of experience. Emmen (2015) mentions the importance of taking the experienced quality of computer assistance into account in future research.

Logg et al. (2019) state that it is important to take the domain of the situation where decision have to be made into account. For instance whether the decision has to do with taste, or with a more objective estimation. They also found that when there is a known historical use by a large amount of people, this will yield in a higher level of acceptance. Therefore the known historical use will be included as a factor in our conceptual framework which we will further discuss in chapter III.

Social influence

In a research by Alexander, Blinder and Zak (2018) it was examined how algorithm adoption, neurophysiological activity and task performance are influenced by the information about algorithms. Their research covered four experimental conditions with different background information about an algorithm. Participants could buy an algorithmic aid to solve a maze puzzle. Participants received – depending on their experimental condition - information about the algorithm related to accuracy (75% of the advice given is accurate), low social proof (54% of the other people use the algorithm) or high social proof (70% of the other people use the algorithm). In the control condition no information was given about the algorithm. Within the two social reference conditions the algorithm was bought more often (26.67% and 25%) compared to the accuracy (18.8%) and control condition (12.5%). Their key finding was that social proof is the most effective way to persuade people to use algorithmic aid, but that the amount (percentage) of social reference does not matter much (Alexander et al., 2018). Therefore we will also add social reference as a factor to the conceptual model, which will be further discussed in Chapter III. Another noticeable finding of Alexander et al., (2018) was that in the condition where no information was given about the algorithm, participants did not check the quality of the advice, which suffered their performance.

The black box

When using algorithms, people are often unaware how the model exactly derives its conclusions, this is the so called *black box*. Consequently, it is important to understand people's default interpretations of what an algorithmic judgment conveys (Logg et al., 2019). Therefore one needs to make sure subjects have the same base level knowledge. If not, this might affect the level of trust one has towards it.

C. Authority

Weber (as cited in Lustig, 2015) defines authority as leadership that is perceived as legitimate and without coercion. Obedience is complying to something that you are told to do by an authority figure. Therefore the term granting authority within this research will be used to describe the willingness to act in accordance with a suggestions given by an agent, measurable by the decision to use the aid without objection. In this section we will first discuss influential historical literature related to authority, after which we will focus on the notion of authority within a human-agent interaction.

Milgram's obedience studies

More than fifty years ago Milgram shocked the world with his obedience studies, showing that ordinary people are able to harm other human beings under the pressure of a scientific authority (Milgram, 1963). Milgram was inspired by the Asch conformity study, where people showed conformity to a group of peers even when they knew the peers were wrong (Asch, 1956). In Milgram's view, conformity is more important regarding issues related to the foundations of social life (Meeus & Raaijmakers, 1995). However, there is some skepticism towards the outcomes of Milgram's experiments. First, such extreme obedience is most likely unique to the laboratory setting of the research. Second, giving electric shocks to people, has not much relation to tasks people have to carry out in their everyday life (Meeus & Raaijkmakers, 1995).

In 1982 and 1983 Meeus and Raaijmakers (1995) replicated to a certain extent Milgram's experiments in "modern society" under a series of experiments called 'the Utrecht studies'. The main difference to Milgram's study (1963) was the type of task during these studies. Meeus and Raaijkmakers (1995) proposed the importance of *mediated violence*, where subjects only indirectly observe negative consequences of their actions. During their experiments on *administrative obedience* the subjects were instructed to make negative remarks about someone's test performance while administrating a test for a job. The obedience of the participants was extremely high. Meeus and Raaijmakers (1995) concluded that psychological-administrative violence can be assumed as a normal social

circumstance in modern Western societies, and that therefore experiments related to this domain are very important.

Authority during human-agent interaction

To what extent do humans obey to artificial agents? Geiskkovitisch, Cormier, Seo and Young (2016), inspired by the earlier mentioned Milgram research, investigated the dynamics and level of obedience during human-robot interaction. Their goal was to find out how people respond to different designs of robots and whether this affected the degree of authority given to the robots. Furthermore they wanted to know whether the level of authority was correlated to the level of obedience (Geiskkovitisch et al., 2016). During 'a very cumbersome and boring task', they tested at which point subjects would want to quit the experiment, and compared these results between a human experimenter and various shapes of robot experimenters. While much more participants (86%) obeyed a human experimenter than they obeyed an autonomous robot (45%), the level of embodiment of the robot (human-shaped, disc-shaped and computer server shaped) did not cause significant differences between the level of obedience. Subjects who attributed authority to the robot protested much earlier (M = 22.85 min) than the group that did not see the robot as an authority (M = 48.73 min). This discrepancy can be explained because the participants believed authoritative robots had the possibility to alter the experiment. The participants told the researchers that they obeyed the robot as they believed that qualified researchers programmed it (Geiskkovitisch et al., 2016). Lustig (2015) also stated that an important factor for trusting an agent, is whether users feel that they can trust the developers, other users, regulators, and anyone else influencing the algorithm (Lustig, 2015). Due to these findings, we will take 'qualification of developers' also into account as a factor for granting authority to agents. We will add it to our conceptual model that is further discussed in chapter III.

Liu et al. (2008) tried to uncover the social psychology of persuasion in human-computer interaction**.** During their experimental design, Liu et al. (2008) asked participants to accept or reject update announcements, depending on the amount of positive or negative attributes they entailed. They designed the task in three phases. During the first phase collaboration and trust was built, by only giving correct aid. During the second phase misleading suggestions were incorporated and during the third phase they tried to restore the trust by giving correct advice. The degree of anthropomorphism did not seem to affect the degree of conformity, which is similar to the outcomes of Geiskkovitisch et al. (2016). Neither did the level of persuasiveness seem to affect the degree of conformity. Rather, Liu et al. (2008) demonstrated that advice should be implicit and subtle in order to facilitate conformity. When messages are too salient, by putting too much emphasis on the agent instead of the human as the central actor, counter intentional effects can occur such as psychological reactance (Liu et al., 2008). Psychological reactance is the human unwillingness to accept external authority in order to retain their autonomy during decision-making. To summarize the findings of Liu et al., (2008) consistent and sophisticated communicative cues achieve a higher level of granting authority towards agents. Furthermore, human-computer interaction depends on the process of trust buildup, rather than the appropriateness of the decision alone.

D.Trust

Trust plays an important role within collaboration, and as collaborative settings will increasingly involve agents due to technological innovations, it is important to research the level of trust during human-agent interaction (Schwaninger, Fitzpatrick & Weiss, 2019). Trust is defined as the confident expectation that one's vulnerabilities will not be exploited in a situation of risk (Corritore, Kracher & Wiedenbeck, 2003). It is a strategy to cope with the uncertainties inherent in human life; trust makes those uncertainties bearable (Keymolen, 2016). It is not always easy for humans to cope with life's complexity and they have done so by creating artifacts (Plessner as cited by Keymolen, 2016). AI technology is an example of such an artifact. The demand for big data is growing, while the basic belief is that if we can gather enough data and find the right correlations, it will become possible to predict and control our environment. The general tendency is that this will contribute to the prevention of risks.

Communication about trust

New technologies, such as machine learning, are becoming more autonomous and consequently the experienced insecurity of consumers grows (van Belkom, 2017). This development has caused that the role of trust in such technologies becomes more important. Van Belkom (2017) argues that communication about the trustworthiness of the specific technology, can improve the level of trust for consumers.

Cai and Lin (2010) reached similar conclusions in their research on how to tune trust using cognitive cues to enhance human-machine collaboration within a driving setting. Their experiment showed that communication about confidence of an agent significantly affected perception of such agents; they were perceived as more useful and trustworthy when showing self-confidence. More specifically they expressed this level of self-confidence by visual adjustments such as size, shape and brightness and by auditory interface such as duration and loudness and haptic feedback. They theorized that if an intelligent machine honestly communicates its self-confidence, it prevents a user from randomly deciding to believe it or not. Furthermore they describe that variable self-confidence make the communication more human-like and less arrogant by agreeing that it is not always correct. We will therefore also add 'self-confidence of a system' as a factor that may facilitate the granting of authority to our conceptual framework, that will be further discussed in chapter III.

Perception and attitude

In a research conducted by Salem et al. (2015) on the trust of faulty robots, it was found that the performance of a home companion robot did not substantially influence subjects decisions to accept its request or not although it did affect the

subjective perception regarding reliability. Important is to mention that the nature of the task might be influential in this matter. Kizilcec (2016) researched the effects of transparency on trust in algorithmic interfaces. He found that knowing how a system actually functions can both induce positive or negative attitudes towards the level of trust of a system. Kizilcec (2016) states that facilitating trust requires a certain degree of transparency but not too much and not too little. Gulati et al. (2017) mention that a person's incentive and willingness to complete a task also impacts trust.

Relation demographics and trust

In previous studies on the relation of demographics on trust during a human-agent interaction, it was showed that older people have a higher level of trust towards automation than younger people (Alexander et al., 2018). Several lines of evidence within this research, suggest that at the beginning of a task women show less trust in automated systems than men, but once women start to trust a system they show a higher level of trust than their male counterparts. Surprisingly, results from the experiment of Alexander et al. (2018) showed that women adopted the algorithmic aid twice as often as men.

Overall, trust seems to play an important role in the decision to use a technology. Trust is a key factor in determining user acceptance of technologies and technology adoption (Gulati, Sousa & Lamas, 2017). Without trust, a user will proceed with more caution and takes more time to think throughout a task. However it must be taken into account that some researches (Alexander et al., 2018 and Cai and Lin, 2010) measured the percentage of adopted behavior of participants and denominated this as the level of trust towards a system.

E. Decision-making

While Artificial Intelligence is the science of knowledge representation and reasoning (Newell & Simon as cited in Pomerol, 1996), human decision-making is a cognitive process that also entails reasoning from the known to predicted outcomes. During decision-making processes, humans tend to experience cognitive biases. For instance, the notion of reactance when feeling that their level of autonomy is being threatened resulting in regaining control by showing opposite behavior (Agrawel & Williams, 2017). According to *the heuristic model of persuasion*, people often follow simple decision-making rules that are based on specific cues, such as the likability of the message source, the perceived expertise or social reference (Liu et al., 2008). An example of the latter is when for instance the majority of people make the same decision.

There is an important trade-off that people need to make during their decision to accept or reject an agent's decision. They may doubt the decision if they cannot validate the trustworthiness. On the other hand, humans have a limited information-processing capacity due to the constraints of the working memory. Therefore they may be tempted to over-rely such agent decisions to reduce their own cognitive load (Häubl & Murray, 2001).

III. CONCEPTUAL FRAMEWORK AND HYPOTHESES

We built a conceptual framework on the factors that contribute to the granting of authority within a human-agent interaction by combining the findings of the research presented in the previous chapter. This model on 'factors influencing granting authority to artificial agents' is displayed in Figure 2. After discussing the framework we will formulate the hypothesis on our research question.

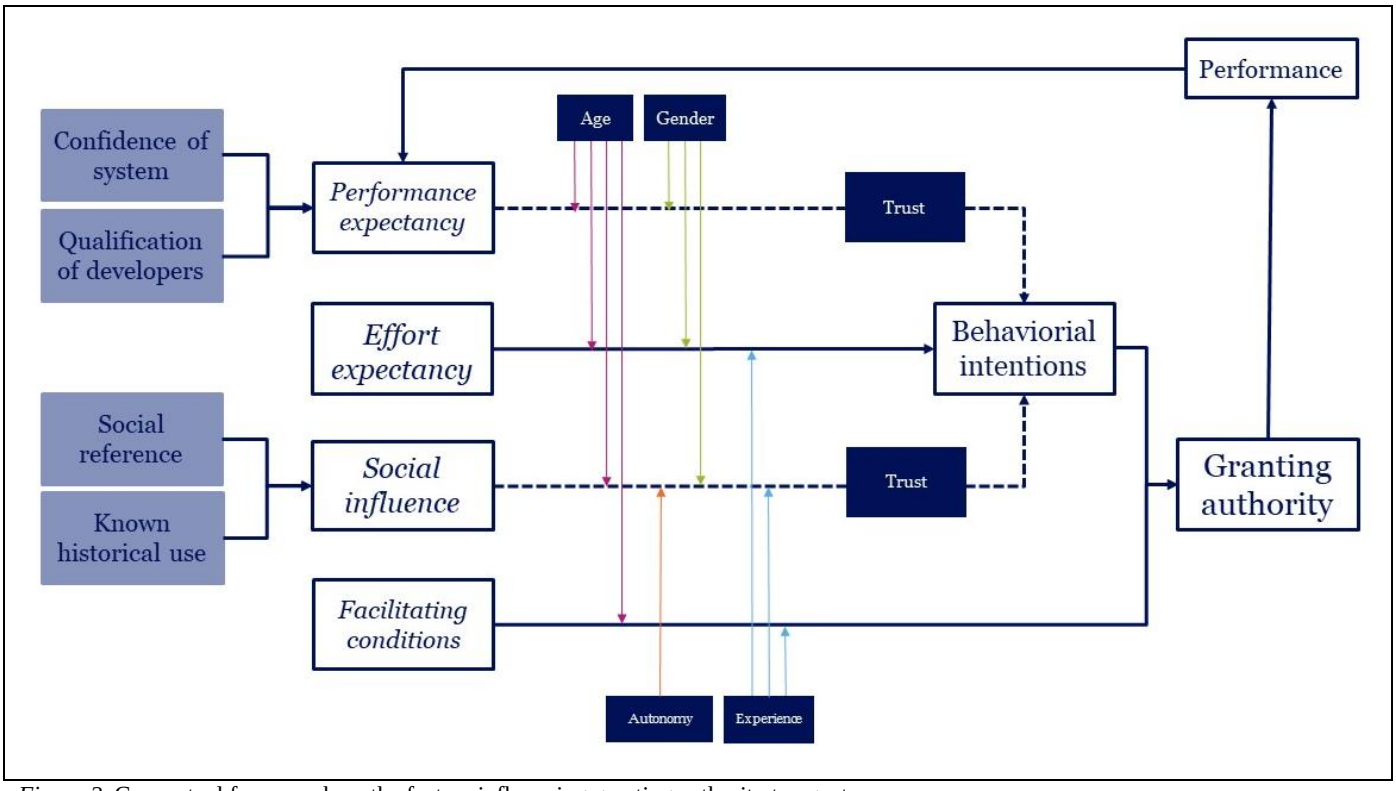

Figure 2. Conceptual framework on the factors influencing granting authority to agents.

A. Conceptual framework

As a foundation for the conceptual model, we used the UTAUT framework established by Venkatesh et al. (2003) that is portrayed in Figure 1. We adjusted and extended this framework based on findings from the literature on adoption of algorithmic advice and obedience to agents to understand what is needed for agents to receive authority.

Within this research we defined giving authority to a system as the decision to confirm with its suggestions. Therefore we substituted the variable 'use behavior' from the UTAUT framework for 'granting authority'. Liu et al. (2008) showed that humans want to retain autonomy, and that a low level of control can result in psychological reactance, which will negatively affect the behavioral intention. Therefore we substituted the need for voluntariness as portrayed in the UTAUT model in Figure 1 for the notion of human autonomy.

We identified four different factors that may influence performance expectancy and social influence, and therefore the authority granted to an (algorithmic) agent: *the self-confidence of a system*, information about the *qualification of the researchers*, *social reference* related to the usage and its *known historical use*. We propose that the factors self-confidence and qualification of developers influence the performance expectancy, while social reference and known historical use affect the social influence.

Correct performance will lead to a trust buildup which will increase performance expectancy in case of an iteration (Liu et al., 2008). Based on findings of Alexander et al. (2018) and the original UTAUT model, gender, age, autonomy and experience are expected to have a moderating role on the granting of authority. We included trust as a mediator in the model, as we found that trust and deciding to accept an agents suggestion are highly inter-linked.

B. Hypothesis

The question we want to answer in this research is*: 'What factors influence granting authority to algorithmic aid during a decision task?'* The part of the conceptual framework which we will explicitly address during this research in order to answer the research question, is displayed in Figure 3. The different experimental conditions are marked blue. We want to examine to what extent each of the defined factors contributes to the granting of authority to an algorithm during a decision task.

From the literature research we expect social reference to have the biggest positive impact on reliance of a system and therefore granting authority to an algorithm. Next, we expect that information about the qualification of developers will have a major effect. Known historical use of a system is expected to have a smaller effect than social reference and qualification of developers, as not much literature mentions this factor. In this research we examine the effect of self-confidence of an agent by providing information about the truthful accuracy of the algorithm. We expect this notion of self-confidence of an agent to have the lowest impact on the level of granting authority to algorithms. Although Cai and Lin (2010) found that systems showing a certain degree of self-confidence are perceived as more trustworthy, people are less likely to use an algorithm after learning about its imperfection (Dietvorst et al., 2014). Also, Alexander et al. (2018) did not find a significant increase in algorithmic adoption in a setting where accuracy information was provided. We expect the control condition, that does not provide any information about the algorithm, to be least successful for adopting algorithmic aid and even expect people in this condition to have a lower overall performance compared to the other conditions, as research of Alexander et al. (2018) showed. Furthermore we hypothesize that the amount of trust is highly correlated with the amount of authority granted. Related to demographic information, we expect man and younger people to have a lower algorithm adoption compared to females and older subjects, as Alexander et al., (2018) mentioned.

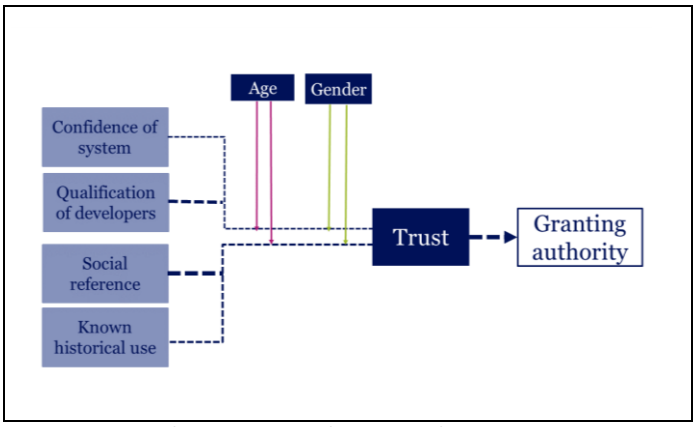

Figure 3. Part of the conceptual framework from Figure 2 that will be addressed during this research. The relative line-sizes show the hypothesized effect size of each factor.

IV. Method

For this research we developed a task called 'The Dismiss Decision'. The Dismiss Decision was inspired by the interactive webpage 'Survival of the Best Fit'; an open source project to teach people more on the dangers and possible biases of using algorithms during the recruitment process (SOBF, 2019). In this section we will further elaborate on the type of task, the design, participants and procedure of the experiment. Furthermore we will discuss the analyses needed to answer the research question.

A. The decision task

In light of the validity of the research it was important that the experiment could be a hypothetical real-life situation. Therefore it was imperative to decide on a task where currently algorithms are being used for. We learned from the literature that a task addressing administrative violence is a possible way to test the level of obedience (Meeus & Raaijkmakers, 1995). This resulted in a task where we addressed decision-making within the field of firing people. During The Dismiss Decision, subjects were told they needed to fire co-workers. Which employee did they consider the weakest, taken into account their *skills, productivity, ambition and experience*? During the experimental

conditions, participants could choose to accept or reject advice from an algorithmic recommendation system. Logg et al., (2019) stressed that the type of task is very influential during experiments related to human-agent interaction. We facilitated that the accuracy of the decisions that participants made was objectively measurable, by informing that the four characteristics about the to-be-fired employees contributed evenly to their functioning. Thus, participants made the correct decision if they chose to fire the employee with the lowest overall scores on the displayed amount of skills, productivity, ambition and experience.

B. Design

The design of the experiment was fairly complex. Therefore we will first elaborate on the different (experimental) conditions. Next we will explain the different phases within the task.

Conditions

The Dismiss Decision was built in Qualtrics and consisted of six different conditions: two control conditions and four experimental conditions. The experimental conditions facilitated the different factors we hypothesized to influence the level of granting authority as presented in Figure 2 and Figure 3. We used a *between-group* design, where all subjects were randomly assigned to one of the following six conditions.

- **Cb0:** *baseline control condition.* Subjects received no algorithmic aid. We used this condition to validate the complexity of the task.
- **C0:** *control condition*. No further explanation about the algorithm was provided.
- **C1:** *self-confidence of agent*. The notion of confidence is used to describe a state of being certain that a hypothesis is correct (Cai and Lin, 2010). Therefore, we facilitated this condition by providing the level of accuracy of the algorithm. During this experiment, the algorithm is 70% accurate.
- **C2:** *known historical use*. Within this condition we provided the message that the algorithm had already been in use by the company for 1.5 years. This amount of time was based on a news article about someone that was fired by a machine, back in June 2018 (Wakefield, 2018).
- **C3:** *social reference.* In this condition a message informed the participants that 70% of the other participants used the algorithm. The percentage is based on the actual percentage of answers correctly provided by the algorithm.
- **C4:** *qualification of developers.* Subjects within this condition received the message that very successful and widely accredited developers built the algorithm.

The participants received 12 tasks (excluding one practice question) during which they had to decide which employee to dismiss. During 10 of these tasks (in condition C0 – C4), we presented an algorithmic advice. We made sure to provide the algorithmic advice implicit and subtle in accordance to Liu et al. (2008) with the words *"it is recommended to fire …."*, in order to prevent discomfort and possible psychological reactance that may affected the level of obedience.

Retaining a minimum level of autonomy is important for a user. To facilitate this, subjects always had the choice to either accept or reject the given aid. We learned from our literature research that it is best to not immediately give feedback about errors of algorithms, as that can immediately affect a user's confidence towards a system (Dietvorst et al., 2014). Therefore the participants did not receive any feedback on their dismiss decisions and only received their overall score at the very end of the experiment. We did not explain how the algorithm exactly drove to its conclusions, since we wanted to keep the level of transparency stable between conditions. Altering transparency can have a big effect on the level of trust according to Kizilcec (2016).

Phases

Within The Dismiss Decision we distinguished three different phases as displayed in Figure 4. A period that facilitates trust buildup is important during a human-agent interaction, as shown by Liu et al (2008). We varied information about the algorithm between the different experimental conditions during each phase transition. Between phase 1 and phase 2, we also provided general information about what an algorithm is to facilitate a baseline level of knowledge among all participants, in accordance with Logg et al. (2019).

- **Phase 1:** Period to practice the task
- Information about what an algorithm is (C0-C4) and
- Experimental information about the algorithm (C1-C4)
- **Phase 2:** Period that facilitates trust buildup
- Repetition of experimental information about the algorithm (C1-C4)
- **Phase 3:** Both correct and misleading suggestions to facilitate possible trust disruption

The amount of time available to make each decision, decreased during each new phase. The time available was 15 seconds during phase 1, 12 seconds during phase 2 and 10 seconds during phase 3. These time limits were the same across conditions to make sure this would not affect the outcomes among conditions. The decision to decrease time, and therefore

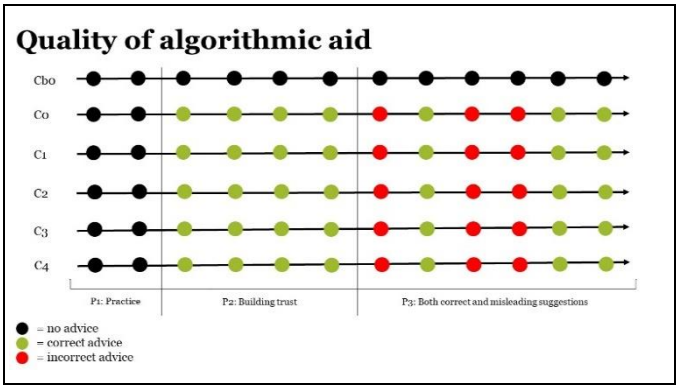

Figure 4. Set-up of the different phases and quality of aid within the different conditions.

possibly triggering granting authority to the algorithm while complexity increases, was inspired by the Surivival of the Best Fit game (SOBF, 2019). In SOBF they explain the need for algorithmic aid as a way to facilitate time reduction and therefore costs for a company, which is a plausible reason for companies to actually use algorithms. We expected that if we would have given participants enough time to always double-check the algorithmic aid and its proposed decision, it would have given them less incentive to blindly follow the algorithm.

Before conducting the experiment we tested The Dismiss Decision with 5 people to make sure the task was understandable and to verify that the set time limits were doable, though still challenging.

C. Participants

The survey was distributed through personal (social) networks such as 'LinkedIn', 'Instagram', 'Facebook' and 'Whatsapp groups'. We also shared The Dismiss Decision through the external networks 'surveywap.io' and 'surveycircle.com'. Participants could make the test anywhere where they had access to internet, but we recommended to do so in a quiet area where they did not get disturbed. Furthermore participants were advised to take the test from a desktop, rather than a mobile phone as the screen resolution would be better. Between 20 December 2019 and 09 January 2020, 221 people completed the survey. Their age ranged from 18 to 60 years (age M = 26.0, SD = 6.98). More females (74.2%) than males (24.9%) took part in the experiment. A complete overview of the demographics can be found in Table A1 in the appendix.

Participants needed to sign for informed consent, before they were able to start with the task. In the consent form participants were informed that the person with the most efficient (= most accurate and fast) test scores, would win €30. To make sure the same subjects did not participate more than once, we disabled people from retaking the survey through a setting in Qualtrics.

D. Procedure

After signing the informed consent, participants received explanation about the to be taken dismiss decisions. Based on a number of characteristics, they needed to decide who was the weakest employee of the department. They received explicit information that the company they worked for valued skills, productivity, ambition and experience equally. These specific characteristics were loosely based on the characteristics used within SOBF (2019). Next, they were given an example task to validate their understanding. This example task is displayed at Figure 5. Participants were not able to continue to the next screen, before entering the correct answer (D) within the text box to make sure they understood what was expected of them. It was again stressed that the company valued the variables skills, productivity, ambition and experience equally. Furthermore it was emphasized that it was important to make the decisions as correct and fast as possible. To further induce incentive, it was again mentioned that a prize of €30 would be rewarded to the person making the best and fastest decisions. It was also explained that, contrary to the example question, they would not receive feedback on the correctness of their decisions. After the example task and provision of information, the participants started with phase 1. To enhance pressure, a timer was displayed during each of the 12 dismiss decisions.

During phase 1 participants had to make two dismiss decisions for which they received 15 seconds each to complete. After phase 1, the participants were told that they needed to make the decisions faster and would only receive 12 seconds for each decision. The experimental conditions received condition-specific information regarding the algorithm. Additionally, to facilitate baseline knowledge about algorithms, the conditions C0 – C5 received the following information: *"An algorithm is a process to be followed in problem-solving operations or calculations. Some algorithms can learn from data and independently find patterns in data to build models and give predictions."*

Phase 2 consisted of four dismiss decisions, during which the algorithm provided gave the correct answer in order to facilitate trust buildup. The participants could choose to accept or reject the aid. An example can be seen in Figure 6. In case of rejecting the aid, they were asked to provide the number of the employee they decided themselves to dismiss. After phase 2 the participants were again told that they needed to make their decisions faster and would only receive 10 seconds per decision. Furthermore, the background information about the algorithms was once again provided within the experimental conditions.

During phase 3 the information displayed was similar to phase 2, however the algorithm sometimes gave the wrong suggestions as can be seen in Figure 4. During phase 3, the participants needed to make six dismiss decisions. For the baseline control condition (Cb0) the procedure was the same, although the participants would not receive information about algorithms nor algorithmic aid.

After The Dismiss Decision, we asked the participants a few explorative questions on a 7-point Likert scale (from 1 = "very high" to 7 = "very low"):

- **Related to the task:** level of complexity, experienced time pressure, perceived autonomy and incentive to make the correct decisions.
- **Related to the algorithm:** perceived quality, amount of trust and authority.

Figure 5. Example task at the beginning of the experiment

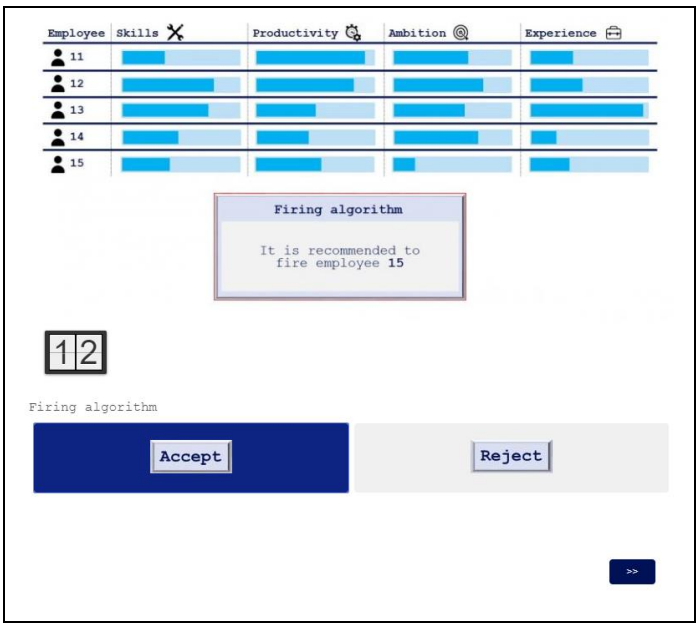

Figure 6. Dismiss decision of phase 2.

The exact questions are presented in Figures B1 and B2 in the appendix. For the baseline control condition, no questions about the algorithm were asked. Following these questions, we asked for demographic information in relation to age, gender, nationality, highest level of completed education and current employment status. Optionally, participants could also answer the question what made them decide to (not) trust the algorithm and leave additional comments or remarks about the test.

E. Analysis

We will perform multiple analyses on the data obtained with IBM SPSS software. First, we will validate the nature of the task. Next, we will answer the research question by testing the relation of each of the variables we included in the conceptual framework displayed in Figure 3 on the amount of algorithmic authority.

Assessing The Dismiss Decision

We will explore whether The Dismiss Decision was doable without any sort of aid and if the complexity increased as intended. Therefore, we will measure the percentage of correct decisions per task and per different phase within the baseline control condition. We will also test for potential differences on the mean outcomes of the baseline control condition (cb0) and the control condition (c0). We will compare the means between the two conditions by conducting *independent sample t-tests* on the scores of the correct decisions and the scores on the answers to the task-related questions.

Factors influencing algorithmic authority

We conducted this experiment to investigate to what extent each of the found variables in the literature play a role within granting authority to algorithmic aid during a decision task. As described before, we defined the level of granted authority toward an algorithm, as the acceptance rate of such a system. Therefore, we will compute the amount of '*algorithmic reliance'* per participant. Algorithmic reliance will be calculated by the amount of (accepted aid/ amount of offered aid)*100, and portrays the percentage of accepted algorithm aid, independent from the correctness of the advice. Within this research, the score on algorithmic reliance is perceived as the amount of authority provided.

To analyze what factors influence the degree of algorithmic reliance, we will conduct multiple *one-way analysis of variances* (ANOVA's). An ANOVA compares whether scores between multiple independent groups differ. We will test whether the amount of algorithmic reliance, and thus authority granted, significantly differs between the experimental conditions. We will also test if there are differences between the conditions on the amount of wrongly accepted decision aid. To explore possible underlying reasons for differences between algorithmic reliance between the conditions, we will also conduct a one-way ANOVA to compare the answers of the exploratory questions between conditions. Furthermore we will examine the correlation between trust and the amount of algorithmic reliance with a *linear regression analysis*. With another regression analysis we will test for a possible correlation between age and the amount of algorithmic reliance. With an independent sample t-test we will explore whether there is a potential effect of gender on the amount of algorithmic reliance.

V. Results

In this section we describe the outcomes of the statistical analyses. First, we investigated the complexity of the decision task and compared the outcomes between the control conditions. Next, the differences between the experimental conditions were thoroughly examined. We also investigated the correlations of the variables trust and perceived complexity on the dependent variable amount of algorithmic reliance. Furthermore we explored the effects of the demographic variables gender and age on the amount of algorithmic reliance. In case of statistical significant results, the *p* values in the tables are accompanied by asterisks, indicating a significance on the level of <.05 (*****), <.001 (******) or <.000(*******). At the end of this section we present key insights provided by participants and a summary of the findings.

In Table 1 an overview is provided of how the different conditions are displayed in the SPSS output. This table also includes the amount of participants per condition (*n*).

Table 1. *SPSS output and amount of participants per condition*

Condition	**Experiment**	**Number in SPSS output**	***n***
Cb0	Baseline control condition (no aid provided)	1	38
C0	Control	2	38
C1	Self-confidence	3	36
C2	Historical use	4	35
C3	Social reference	5	36
C4	Qualification of developers	6	38

For our analysis we reversed the code scale items of the explorative questions (Figures B1 and B2 in the appendix), which resulted in a scale ranging from 1 = "very low" to 7 = "very high".

A. Assessing The Dismiss Decision

We examined whether The Dismiss Decision was doable without any sort of aid and if the complexity increased as intended. Furthermore we compared the amount of correct decisions, and explored differences in the task-related questions between the baseline control condition (without algorithmic aid) and control condition (with algorithmic aid).

Task complexity

To explore the complexity of the task without algorithmic aid, we examined the decisions of all the subjects in the baseline control condition (Cb0, n = 38). During every task subjects had to choose who to dismiss between five employees, but one decision was objectively the best. In Figure 7 a plot on the average percentages of correct decisions within each phase is presented. The average amount of correct decisions of each of the 12 tasks, can be found in Table C1 of the appendix. There is a trend that the percentage of the average amount of correct decisions decreases per phase, indicating that the complexity of the task increased over time. This difference seems much stronger between phase 1 and 2 than between phase 2 and 3. We observed that almost all participants chose to dismiss the correct person during task 11 in phase 3 (92.1%) which therefore can be considered as a relatively easy task.

Overall scores

The maximum score participants could receive during the experiment was 24: one received 2 points per correct decision. To compare our two control groups we performed a t-test. We validated the assumptions of normality, homogeneity of variance and independence needed to do so. After removal of one outlier within condition C0, the independent sample t-test showed that the difference in total score between the baseline control condition (M =18.16, SD = 3.33) and control condition (M = 18.54, SD = 3.16) was not significant, t(73) = -.510, p = .612. This indicates that receiving algorithmic aid or not, did not affect the accuracy of the respondents.

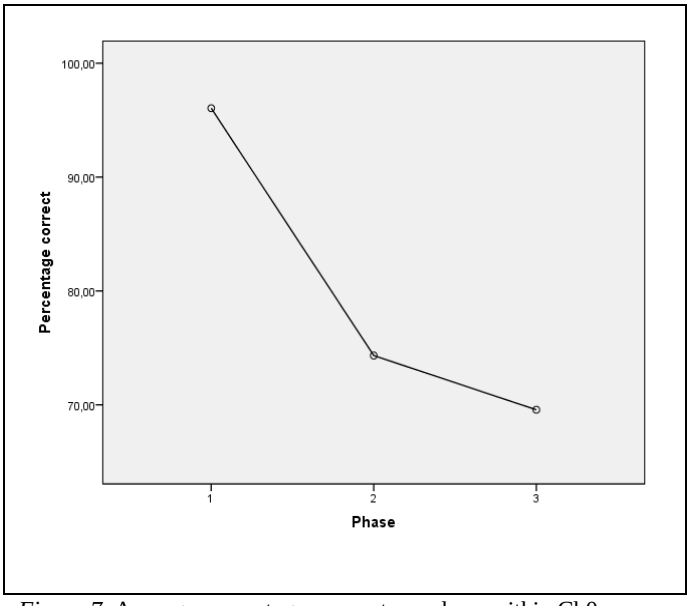

Figure 7. Average percentages correct per phase within Cb0.

Explorative questions

We examined for differences between the answers on the task-related questions between Cb0 and C0 with an independent sample t-tests. None of the scores showed a significant difference, although the perceived autonomy was marginally lower during the algorithmic condition (M = 4.24, SD = 1.55) in comparison with the baseline control condition (M = 4.89, SD = 1.64) with t(74) = 1.797, p = .076. An overview of the means, standard deviations and p values provided by the independent sample t-tests can be found in Table C2 in the appendix.

B. Experimental conditions

In this section we describe the outcomes of the analyses we conducted to compare the experimental conditions. First, we discuss the amount of algorithmic reliance per condition. Next, we examine the amount of wrongly accepted decision aid per condition. Lastly, the scores on the explorative questions are compared between conditions.

Amount of algorithmic reliance per condition

The value of the algorithmic reliance is the percentage of offered help accepted. We calculated the amount of algorithmic reliance per subject, by dividing the amount of accepted aid by the amount of offered aid and multiplying this value by hundred. We will use this value to define how much authority was granted to the algorithm. First, an outlier analysis was performed, by creating boxplots of the algorithmic reliance per condition. An outlier is defined as minimally 1.5-3 box heights from the box. The boxplots generated for the outlier analysis are displayed in Figure C7 in the appendix. This analysis revealed seven outliers within condition 3 of which two were even marked as extreme scores (> 3 box heights from box). One outlier was found within condition 4 and one in condition 6. Because the outliers seemed influential for our data, we decided to remove all of them before conducting further analysis on the amount of algorithmic reliance. The histograms of the raw data of the amount of algorithmic reliance per condition are displayed in Figures C1 – C6 in the appendix.

We checked if we met the assumptions for performing an ANOVA. Because n > 15 in every cell, we concluded that our F is robust, indicating that we met the assumption of normality. Because the group sizes were approximately equal (38/28 < 1.5), even after removal of the outliers, the assumption homogeneity of variance was also met. A one way-ANOVA with dependent variable algorithmic- reliance and independent variable condition showed a significant difference between the different conditions, F(4,168) = 3.683, p = .007 < .01.

Table 2. *Post hoc Tukey HSD. Dep. variable: algorithmic_reliance.*

Condition	Mean Difference	Standard Error	Sign. (*p*)
2 vs 5	11.60	3.37	**.016***
4 vs 5	11.10	3.46	**.014***
6 vs 5	10.31	3.39	**.023***

Post hoc multiple comparison using the Tukey HSD test showed that all the experimental conditions apart from condition 3, significantly differed from condition 5 (Table 2). In Figure 8 the means of algorithmic reliance per condition, including their 95% Confidence Intervals (CI), are presented. These results indicate that the algorithmic reliance is significantly different between the experimental groups. Subjects in condition 5 (social reference) accepted significantly less aid than subjects in condition 2 (control), condition 4 (historical use) and condition 6 (qualification of developers).

Amount of wrongly accepted decision aid per condition

Outlier analysis did not show any outliers. One-way ANOVA was not significant (p = .793), which indicated that there were no major differences in the amount of wrongly accepted decision aid between conditions.

Overall performance and explorative questions

To test for the differences on overall performance and answers to the explorative questions, one-way ANOVA's were conducted. A complete overview of the output of these analysis is displayed in Table C3 in the appendix. For the scores on overall performance, we removed six outliers from the data. The conditions did not significantly affect the total amount of correct decisions taken (p = .181), which indicates that accuracy was not influenced by the information provided about the algorithm. For the analysis on the explorative questions, we included the data of all subjects while Likert-scale scores do not show representative outlier behavior due to the clear floor (1) and ceiling (7) of the scores. The level of perceived complexity differed significantly among conditions, F(4, 178) = 3.374, p = .011 < .05. Post hoc Tukey HSD analyses showed that these differences were found between condition 3 (M = 4.00, SD = 1.45) and condition 5 (M = 5.06, SD = 1.12) with p = .004, < .01. This indicates that research participants experienced the complexity of the task significant easier in condition 3 (self-confidence) than in condition 5 (social reference).

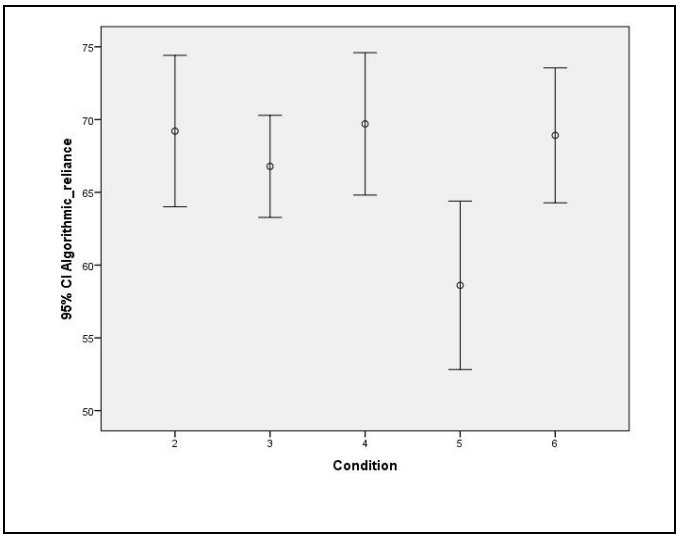

Figure 8. Mean amounts of algorithmic reliance and CI per condition.

C. Correlations on algorithmic-authority

We tested the correlation between the level of trust and the amount of algorithmic reliance. Because we found differences on the perceived complexity and perceived quality of the aid between conditions, we also performed two linear regression analyses to explore the correlation of the perceived level of complexity and the amount of algorithmic reliance and the correlation of the perceived quality of the aid and the amount of algorithmic reliance.

Correlation trust and algorithmic reliance

A linear regression showed a significant positive correlation (p < .000) of .411 between the level of trust and amount of algorithmic reliance, which is an effect size of medium to large (Table 3). This indicates that a higher level of trust, resulted in a higher degree of acceptance of algorithmic advice.

Correlation perceived complexity and algorithmic reliance

We found a significant negative correlation (p = .004 < .001) of -.209, indicating a low to moderate effect size between the perceived complexity and amount of algorithmic reliance (Table 3). This shows that higher perceived task complexity, indicates a lower amount of algorithmic reliance.

Correlation perceived quality and algorithmic reliance

We found a highly significant positive correlation (p < .000) of -.209, indicating a moderate effect size between the perceived quality of the aid and the amount of algorithmic reliance (Table 3). This demonstrates that a higher perceived quality of the algorithm, results in a higher amount of algorithmic reliance.

Table 3. *Output linear regression analyses of level of trust, perceived task complexity and perceived quality of algorithm. Dependent variable: algorithmic reliance.*

	B	St. Error	t	Sign.(*p*)	Part(ial) Correlation
Constant	41.48	4.16	9.97	**.000*** **	
Trust	5.27	.87	6.07	**.000*** **	.411
Constant	78.50	4.61	17.04	**.000*** **	
Perceived Complexity	-2.82	.98	-2.88	**.004** **	-.209
Constant	43.92	5.16	8.52	**.000*** **	
Perceived quality	4.67	1.07	4.36	**.000*** **	.308

D. Demographic variables and algorithmic reliance

We analyzed the influence of age on algorithmic reliance. We also re-performed earlier analysis on a more homogenous age group. Furthermore we investigated the influence of gender on the amount of algorithmic reliance.

Effect of age on the amount of algorithmic reliance

We conducted a linear regression analysis on the correlation between age and the amount of algorithmic reliance. This analysis showed a non-significant result with $F(1,179) = 2.328$, $p = .129$. However, a scatterplot did show a negative trend for the amount of algorithmic reliance related to age (Figure C8 in the appendix). Therefore we examined whether a more notable difference could be observed, if we divided the subjects between two age groups. As a cut-off point we chose 26 years, to distinguish more or less between students- and working people. Outlier analysis removed the data of one participant within each group. Assumptions were met and an independent sample-t test showed that subjects of maximum 25 years accepted algorithmic aid significantly more often ($M = 68.57$, $SD = 14.86$), than subjects of 26 years and older ($M = 62.50$, $SD = 17.72$) with $t(177) = 1.416$, $p = .017$, $< .05$. In Figure 9 a plot of the means and the 95% CI of the two groups can be seen.

Homogenous age group

Because of the observed differences between the two age groups, we performed an outlier analysis on the homogeneity of age of all participants. This analysis showed that the participants of 35 and older could be considered as outliers within our data (Figure C9 in the appendix). After removal of the participants of 35 and older, our *n* decreased to 165. We conducted new one-way ANOVA's with the new subset, to test whether a more homogenous age group would impact the earlier found results. The most notable results are summarized in Table 4. These results indicate that within the new subset, the differences on the amount of algorithmic reliance and perceived complexity between the conditions is still statistically significant and comparable to the results of the whole group.

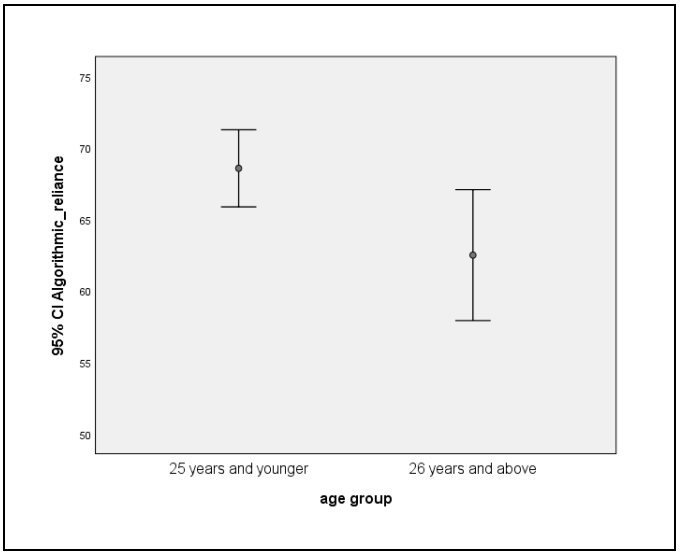

Figure 9. Means plot with 95% CI intervals of algorithmic reliance between the subset <26 and 26+

Table 4. *Output one-way ANOVA's on the differences between conditions for subset age < 35 years.*

		df (between,within)	F	Sign. (*p*)
Algorithmic reliance		4, 155^	3.570	**.008****
Explorative questions				
Task related	Complexity	4, 160	3.333	**.012***
Algorithm related	Trust	4, 160	2.091	**.084**
	Quality	4, 160	3.648	**.007****

^ 5 Outliers were removed

Table 5. *Post hoc Tukey HSD for ANOVA's for subset age < 35 years.*

	Condition	Mean Difference	Standard Error	Sign. (*p*)
Algorithmic reliance	2 vs 5	10.40	3.51	**.029***
	4 vs 5	11.29	3.64	**.019***
	6 vs 5	11.58	3.60	**.014***
Explorative questions				
Complexity	3 vs 5	-1.16	.32	**.004****
Quality	5 vs 6	-.91	.27	**.009****

The differences between the conditions on the perceived trust and quality however have increased. Subjects younger than 35 within condition 6 (qualification of developers) perceived the quality of the aid significantly better ($M = 5.21$, $SD = .98$) than subjects within condition 5 (social reference) ($M = 4.30$, $SD = 1.24$) with $F(4,160) = 3.648$, $p = .007 < .0$. There is a marginal effect of the condition on the perception of trust in the algorithm. The results of the post hoc Tukey HSD's on the subset of participants younger than 35 years are presented in Table 5.

Effect of gender on algorithmic reliance

We tested for differences between gender on the amount of algorithmic reliance, across all participants. Outlier analysis removed scores of three males and of three females. An independent sample t-test between the mean difference in percentage of algorithmic reliance between male participants ($n = 38$) and female participants ($n = 138$) was not significant ($p = .354$).

E. Insights participants

On the question why participants thought they did (not) trust the algorithm, we collected many different responses. The most distinctive answers on why they trusted the algorithm were:

- Stress / time pressure
- I chose to respond quick over accuracy
- It seemed to be mostly right
- I checked it in the beginning and saw it was right

- It felt like the safest choice
- Because it was developed by very qualified developers *(qualification of developers condition)*
- Because they already used it for 1.5 years so that is a good sign to trust it *(historical use condition).*

The most notable comments of participants who did not trust the algorithm were:

- I want to be able to decide for myself
- I saw the algorithm made a mistake, after which I lost trust in it
- Because of the consequences of the action. People are getting fired, it shouldn't be an easy decision and based on machines making it.
- Could be outdated *(historical use condition)*
- Because it was only 70% correct *(confidence condition)*
- I don't know who programmed it *(control condition)*

F. Summary of results

We provide a summary of the most important findings of the analysis we conducted. First, we address general findings related to the task we developed. Second, we discuss the factors that we found to influence the granting of authority during this experiment.

The Dismiss Decision

Without algorithmic advice the average score was 18.16 on a maximum of 24, meaning that subjects made the correct decisions in about 9 out of 12 tasks. This indicates that the task we developed was not very easy to conduct autonomously. Overall, we observed a decreasing amount of correct decisions over the different phases, although this observation was mainly visible between phase 1 and phase 2. Therefore we can conclude that the complexity of the task grew between phase 1 and phase 2. These results indicate that validation of the algorithmic advice during the experimental conditions was not always possible. We also found that providing algorithmic aid did not influence the overall scores of the participants. Furthermore, the perceived autonomy was marginally lower during the algorithmic condition. This indicates that providing advice may have negatively affected the feeling of autonomy.

What factors influenced granting authority to algorithmic aid during The Dismiss Decision?

From our literature research we found that the factors *self-confidence, known historical use, social reference and qualification of developers* play a role in granting authority to agents. We expected, as portrayed in Figure 3, that social reference would be the most important factor, followed by the qualification of developers. We expected a smaller effect of known historical use and the lowest, but still observable, effect of the factor self-confidence of a system. By comparing the amount of algorithmic reliance between the different conditions, we only found significant differences of social reference (condition 5) in relation to the other conditions. Therefore we conclude that social reference played the most important role on the amount of authority granted to our algorithmic aid. Contrary to the hypotheses, the effect of providing information about social reference was negative. We furthermore found that the factor social reference increased the amount of perceived complexity.

We expected information about self-confidence (condition 3) of a system to have the least effect of the defined factors. However, our results show that subjects within this condition perceived the complexity of the same task lower than the subjects within the social reference condition. We also found a low to moderate negative correlation between perceived complexity and amount of algorithm adoption. Combining these findings may indicate that the factor self-confidence influences the granting of authority, although not directly.

While we did not observe a direct effect of information about qualification of the developers (condition 6) on the amount of algorithmic reliance, this condition positively influenced the perceived quality of the algorithm for the sub-set of subjects below 35 years. We also found a moderate positive correlation between the perceived quality of the algorithm and the amount of algorithmic reliance. Together these findings might indicate that the factor qualification of developers influences the granting of authority, although not directly. Known historical usage (condition 4) about the algorithm did not show any effect within our analyses therefore we conclude that this factor did not contribute to the granting of authority within the experiment we conducted.

We expected subjects within the control condition (condition 2) to score lowest on the amount of reliance on the algorithmic aid. We also hypothesized that this group would have a lower overall performance compared to the other conditions. These expectations were not found in our analyses.

No significant differences were found between the various conditions on the amount of trust towards the algorithm, which is contrary our hypotheses that trust mediates the effects of the factors on granting authority. In line with our hypothesis, we did find a moderate to strong positive correlation between trust and algorithmic reliance.

Related to demographic information, we expected males and younger people to have a lower level of algorithm adoption compared to females and older subjects. Contrary this hypothesis we found that younger people tend to show a higher amount of algorithmic reliance. Gender did not seem to have a significant effect on the amount of authority granted to an algorithm during The Dismiss Decision.

VI. Discussion

We will first address points of discussion related to the literature research, after which we will take a critical look at The Dismiss Decision. The participants from our study were able to leave remarks and suggestions about the test which we will also discuss in this section. Next, the possible explanations of the findings from the experiment will be discussed, during which we will also look at possible interaction effects of the variables from

the conceptual framework. To conclude, we will address the practical relevance and limitations of this study and provide insights for future work.

A. Literature research

For this literature research and the conceptual model we thereby constructed, we took the UTAUT model as a foundation. Officially the UTAUT model is a framework that was developed through examining the acceptance of technology within an organization. Therefore it is debatable whether it is transferable towards the use of a specific AI technology as we did within this research. Reflecting on this decision, we did experience the UTAUT model as a plausible point of departure to construct a new conceptual framework on the granting of authority during a human-agent interaction.

During our background research, we combined many different types of research. Some literature addressed a specific human – algorithm interaction, while some conveyed a human – robot or human – computer interaction. Consequently all of these studies had a different design and also varied in outcomes. The relation of trust on granting algorithms authority was difficult to place in a conceptual framework, as some studies take trust in an computational aid and using this aid as a synonym.

B. The Dismiss Decision

By developing The Dismiss Decision, we were able to transfer a domain of AI application into a research setting. By providing information on the amount of skills, productivity, ambition and experience of fictive employees, we could measure the accuracy of decisions on an objective scale. However, multiple participants expressed their confusion about the decisions they had to make. Although the instructions multiple times clearly mentioned that the goal was to find the employee that was the 'weakest link', by averaging the four characteristics, some participants mentioned that they 'did not agree' with this task. For instance because personally they believed that ambition is much more important for an employee than experience. This is an important finding that should be taken into account; the task was essential within our setup, and disproval with the design has probably affected the results for these specific participants. We also discovered, through the comments people left, that the task was very stress inducing due to the limited time available per decision. That the task induced a stress response, indicates that people probably had a high incentive to perform well. However, the well-being of subjects should always be highest priority within research. It is also likely that the high time pressure made the task unrealistic, while firing people does not happen (yet) within a few seconds in real life. It is reasonable that people only decided to accept the decision of the algorithm because they did not have any other choice, as they were not fast enough to do make the decisions in time themselves. Another point of discussion related to the design of the task is that the firing algorithm made its first mistake during the first task of phase 3. This might be argued as unfair because up to that point people had more time to double check the advice. Although we saw a trend of increased complexity to make the correct decisions as the tasks evolved, the difficulty between tasks fluctuated. While participants within every condition needed to conduct the same (order of) dismiss decisions, we controlled for this differences in the set-up of this experiment.

C. Factors behind granting authority to algorithms

Analyses of our data showed the controversial outcome that providing social reference had a direct negative effect on the granting of authority within this experiment. This is highly unexpected, while Alexander et al. (2018), who conducted an experiment with a comparable set-up, showed that social reference has a highly positive effect on algorithmic adoption. This was even the case when they provided a lower amount of social reference than the 70% we used. An explanation could be that our task was perceived as highly unrealistic, and consequently participants could not believe that 70% of their peers adopted the aid. Possibly they became extra aware of the possible catch that was hidden in the experiment. Counter this notion, no subject gave an answer to the question why they did (not) trust the algorithmic aid that validates this theory. Rather, subjects within the social reference condition explained they wanted to "just decide for themselves". Nevertheless, this finding does prove that social reference is a very influential mechanism.

Subjects within the experimental condition that addressed self-confidence of the algorithm, showed a significant lower perceived task complexity than subjects within the social reference condition. This can be explained because self-confidence, in our research tested by the level of accuracy, reveals a certain degree of transparency. Such information makes people conscious that the algorithm is not perfect and prepares them to act accordingly: learning that an algorithm is imperfect, provides the insight that the proposed solution should be checked. The perceived task complexity was significantly negative correlated to the amount of algorithmic reliance. If the task was perceived as more complex, participants accepted the proposed decision less often. This is unexpected, as people are tempted to over-rely agents in situations where their working memory is constraint (Häubl & Murray, 2001). Perhaps we find the explanation of this effect in the nature of the task: the decisions involved administering violence towards others. An answer of a participant on the question why he did (not) trust the algorithm, that understates this theory was: *"Because of the consequences of the action. People are getting fired, it shouldn't be an easy decision and based on machines making it."* The counter-intuitive effect of social reference on algorithmic authority can possibly also be explained by the fact that the task administered administrative violence.

Interestingly, information about qualification of the developers resulted in an increased perception of the quality of the algorithm, but only for people younger than 35 years. This fits the hypothesis that age has a moderating effect. An increased perception of quality of the algorithm positively correlated with the level of higher algorithm reliance.

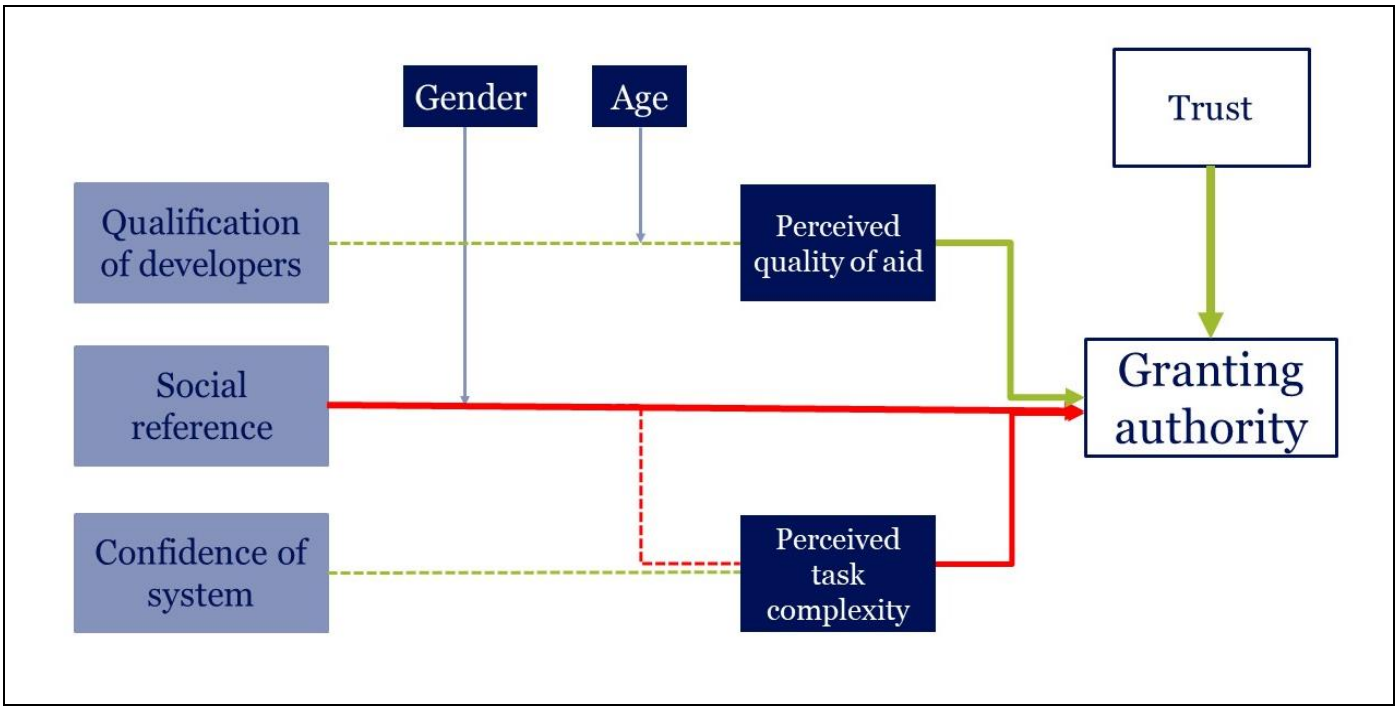

Figure 10. Factors influencing the granting of authority to algorithms during a decision task

We did not find a contributing effect towards our conceptual framework of the variable known historical use. One remark of a participant within this condition was that 'he thought this meant that the model was outdated'. This argument explains why participants did not accept algorithmic aid more within this condition than in the control condition.

A negative trend between age and algorithmic reliance was observed. Only when we divided the subjects in two age groups, this effect was significant. Contrary earlier literature findings, younger people (< 26) accepted the decision of the algorithm significantly more often than older people (> 26). An explanation could be the difference in amount of work experience between the two groups. Possibly more work experience made the task more relatable to a real life setting and the possible consequences of such decisions. It could also be the case that the older group perceived the task of firing the employee that scored lowest on average on a few statistics, as more unfair than the younger participants with less work experience. Although we found a positive correlation between the amount of trust and the granting of authority, the amount of trust was not influenced by the different experimental conditions.

In Figure 10 a final overview of the factors that we found to influence granting authority to algorithms is presented. The thickness of the lines indicate the relative effects. Red lines correspond with the negative relations that we found. Green lines indicate a positive relation and blue lines indicate that we observed an effect but could not verify it with certainty with the data collected within this research. We found social reference to have a direct negative effect on granting authority to an algorithm, although this relation may be (partially) mediated by the perceived task complexity. Perceived quality is found to mediate the effect of information about qualification of the developers. An important remark is that we could not explicitly test the predicted *moderating* effects of age and gender while our independent variables were dichotomous. In relation to the earlier defined framework in Figure 3, we removed the factor of historical use while we did not find any relation to the other variables in the model on the observed data of this condition.

D. Exploration interaction effects

Despite our hypothesis that females accept more algorithmic aid than man, we did not find a direct effect of gender on the amount of algorithmic reliance. This might be due to an interaction effect, meaning that the influence of the experimental condition was different for males and females. We explored the existence of such an effect, by plotting the estimated marginal means (Figure 11). The lines intersect, indicating indeed an interaction effect; it seems that mostly men caused the differences in algorithmic reliance within condition 5 in comparison to the other conditions. Although there is a specific trend visible, we cannot draw a concise conclusion from this plot. The assumption of normality is not met, while there are not enough (>20) data points of both female and male in each condition. Another important finding related to gender and the granting of authority, is that if we exclude all males from the data the effect of the condition on the amount of algorithmic reliance is not significant anymore.

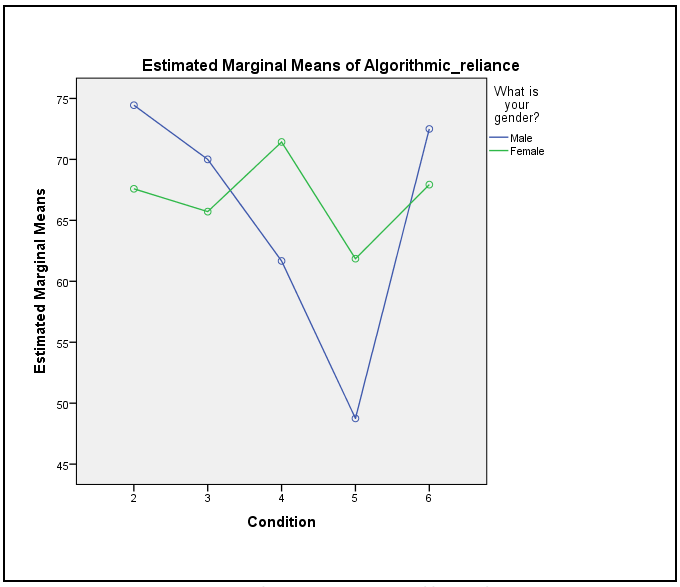

Figure 11. Exploration of an interaction effect of gender and condition on algorithmic reliance.

E. Practical relevance

In the introduction we proposed that the findings from this research may contribute to the knowledge on how artificially intelligent systems should further be developed and designed for interaction. Our findings shed light on the seemingly easy way to manipulate the relinquishing of decision-making to agents, by altering only a small amount of information about the algorithmic decision aid. More specifically we found that social reference is very important in this setting. Although earlier research showed that social reference has a positive effect on the level of trust towards an agent, this research revealed that during a task where violence has to be administrated this effect might work contradictory. The contradictory results of this study in relation to previous research, shows that more research on the topic of algorithmic authority is highly needed.

We also found evidence that varying background information about an algorithm can lead to a higher level of perceived quality of the algorithm or to a perception of decreased task complexity, while in reality the quality and task complexity remain exactly the same. This is useful knowledge that proves it is possible to design human-agent interfaces in such a way that the perceived complexity of a task decreases. Information about the self-confidence of an agent, by providing the level of accuracy of an algorithm, may lead to this result.

F. Limitations

Specifically related to the experimental set-up and design of The Dismiss Decision, it is important to mention a few limitations. Although it was highly recommended to perform the task on a desktop, many participants took the test on their smartphone. This had a negative effect on the readability and caused time delay while the participants needed to scroll on their screen during the tasks. We did not measure which platform was used, thus could not account for the potential differences that this caused. However, because the conditions were randomly assigned, it is likely that every condition entailed similar amount of mobile and desktop users. Another remark that participants addressed, was that they had to explicitly press "next" after accepting or rejecting the algorithm, which was not clear from the start for everyone. This may have caused some time delays. It is also important to mention that the results of this research may not be easy transferable to other settings where algorithmic aid is being used. In the real world, there is an uncountable amount of different scenarios and contexts that influence decision-making (Logg et al., 2019). Nevertheless it is important to gain more insights into to contributing factors to the usage of decision-aid during human resources related task, as artificially intelligent systems play an increasing role within this domain.

G. Future work

Future work related the topic of algorithmic authority should, similar to this research, research tasks in which algorithms currently are being used. In contrast to this research however, it would be better if the task was more realistic. For instance by providing more time or using a decision task where humans are more familiar with. Furthermore it would be interesting to find out if similar results would be obtained during 'more positive' tasks, for instance a task involving decisions on who to hire. We recommend future work to vary the independent variables of our model displayed in Figure 10. For instance by designing an experiment where the degree of qualification of the researchers, the level of confidence or the amount of social reference is tested as ordinal or continuous variable. The findings of such experiments can provide more insights on the mediating and moderating effects involved in granting authority to algorithms and extend our knowledge on how algorithmic advice should (not) be presented in order to increase its authority.

Automated decision-making undermines our own proactive decision-making. Siebert, Kunz and Rold (2019) found that proactive decision-making can improve life satisfaction as mediated by decision satisfaction and self-efficacy. Focusing too strong on a predictable and controllable environment may furthermore hinder the 'societal intelligence and resilience inherent in human life to thrive' (as mentioned by Dewandere in Keymolen, 2016). Therefore another idea for a future research direction that we propose is to investigate if life satisfaction or happiness decreases in situations where decision-making is relinquished to agents. If so, this has huge implications on to the use of artificially intelligent systems.

As a final recommendations for future work we suggest to dedicate more research to the topic of 'Theory of Machine', analogous to the 'Theory of Mind', on how people theorize about the input, processing and value of the output of the information offered by an agent (Logg et al., 2019). We also propose to further explore 'machine behavior'. Machine behavior is the idea that we should observe machines in their natural habitat and develop theories based on those observations in an order of action-stimulus-reaction. Defining the input and output in different settings, might be a way to map what happens inside the mysterious black box of AI.

VII. Conclusion

In this research we examined which factors influence granting authority to algorithms during a decision task. Within literature we found that *self-confidence of an agent, the known historical use, social reference and qualification of developers* are factors that contribute to the level of obedience towards an agent. We developed an experimental task called The Dismiss Decision. Findings of this experiment suggest that information about social usage of an algorithm has a negative influence on the amount of granted algorithmic authority during decisions on which employee to fire. Information about the qualification of the developers positively influenced the perceived quality of the algorithm within younger subjects (< 35 years). Information about the self-confidence of a system resulted in a lower perceived task complexity. Known historical use of an algorithm did not show significant results within our experiment. A low to moderate negative correlation was found between perceived task complexity and algorithm adoption. A moderate positive correlation was found between perceived algorithm quality and the amount of accepting algorithmic advice. Surprisingly, amount of trust was not affected by the different experimental conditions. Age seems to negatively influence the level of granted authority to an agent. Future research should further examine the moderating effect of age and gender and the mediating effects of perceived algorithm quality and task complexity on the amount of algorithm authority.

Acknowledgments

We would like to thank Stichting Toekomstbeeld der Techniek (STT) and specifically Rudy van Belkom, for offering inspiration throughout the research and giving insights into the current and future opportunities and challenges related to artificial intelligence. Furthermore we would like to thank the supervisors Peter van der Putten and Francien Dechesne for their guidance during this project.

References

[1] Alexander, V., Blinder, C., & Zak., P. J. (2018). Why trust an algorithm? Performance, cognition, and neurophysiology. *Computers in Human Behavior*. *89*, 279–288.

[2] Agrawal, S., & Williams, M. (2017). Robot Authority and Human Obedience: A Study of Human Behaviour using a Robot Security Guard. *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 57-58).

[3] Araujo, T., de Vreese, C., Helberger, N., Kruikemeier, S., van Weert, J., Bol., N., … Taylor, L. (2019). *Automated decision-making fairness in an AI-driven world: public perceptions, hopes and concerns.* 1-21.

[4] Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological monographs: General and applied. 70*(9), 1-70.

[5] Bodo, B., Helberger, N., Irion, K., Zuiderveen Borgesius, F., Moller, J., van de Velde, B., … de Vreese, C., (2017). Tackling the algorithmic control crisis-the technical, legal, and ethical challenges of research into algorithmic agents. *Yale JL & Tech*. *19*, 133-179.

[6] Cai, H., & Lin, Y. (2010). Tuning Trust Using Cognitive Cues for Better Human-Machine Collaboration. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 2437-2441).

[7] Corritore, C. L., Kracher, B.,& Wiedenbeck. S. (2003). On-line trust: concepts, evolving themes, a model. *International Journal of Human-Computer Studies. 58*(6), 737–758.

[8] Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. Retrieved from https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

[9] Dietvorst, B. J., Simmons, J. P., & Massey, C. (2014). Algorithmic Aversion: People Erroneously Avoid Algorithms. *Journal of experimental psychology: general*. 1-13.

[10] Dietvorst, B. J., Simmons, J. P., & Massey, C. (2016). Overcoming Algorithm Aversion: People will Use imperfect algorithms if they can (even slightly) modify them . *Management science*. 1155-1170.

[11] Emmen, D. (2015). Checkmate! The Willingness to Accept Computer Aid. *Master's Thesis for the Media Technology programme, Leiden University*. <unpublished>

[12] Geiskkovitsch, D. Y., Cormier, D., Seo, S., & Young, J. (2016). Please Continue, We Need More Data: An Exploration of obedience to robots. *Journal of human-robot interaction*, 82-99.

[13] Gulati, S., Sousa, S., & Lamas, D. (2017). Modelling Trust: An Empirical Assessment. *IFIP Conference on Human-Computer Interaction* (pp. 40-61).

[14] Hale, K. (2018, April 5). Leveraging the Power of AI in Marketing, Now and In the Future. Retrieved from https://towardsdatascience.com/leveraging-the-power-of-ai-in-marketing-now-and-in-the-future-42de905e8274.

[15] Häubl, G., & Murray, K. (2001). Recommending or persuading? The impact of a shopping agent's algorithm on user behavior. M. Wellman & Y. Shoham (Eds.), *Proceedings of the ACM conference on Electronic Commerce* (pp. 163–170).

[16] Keymolen, E. (2016). *Trust on the line: a philosophical exploration of trust in the networked era*. Oisterwijk: Wolf Legal Publisher.

[17] Kizilcec, R.F. (2016). How much information? Effects of transparency on trust in an algorithmic interface. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2390–2395). ACM.

[18] Liu, S., Helftenstein, S., & Wahlstedt, A. (2008). Social psychology of persuasion applied to human-agent interaction. *An interdisciplinary journal on humans in ICT environments*, 123-143.

[19] Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 90-103.

[20] Lustig, C., & Nardi, B (2015) Algorithmic authority: The case of bitcoin. *System Sci. Proceedings of the 48th Hawaii International Conferance* (743–752).

[21] Lustig, C., Pine, K., Nardi, B., Irani, L., Lee, M.K., Nafus, D., & Sandvig., C. (2016). Algorithmic Authority: The Ethics, Politics, and Economics of Algorithms That Interpret, Decide, and Manage. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1057–1062).

[22] Makridakis, S. (2017). The forthcoming artificial intelligence (AI) revolution: Its impact on society and firms. *Futurus. 90*, 46-60.

[23] Meeus, W.H.J., & Raaijmakers, Q.A.W. (1995). Obedience in modern society: the Utrecht studies. *Journal of Social Issues. 51*(3), 155–175.

[24] Milgram, S. (1963). Behavioral study of obedience. *Journal of abnormal and social psychology, 67*, 371-378.

[25] Orwell, G. (1949). *1984.* London: Secker and Warburg.

[26] Pomerol, J. (1997). Artificial intelligence and human decision-making. *European Journal of Operational Research. 99*, 3-25.

[27] Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015). Would You Trust a (Faulty) Robot? *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15.*

[28] Siebert, J., Kunz, R., & Rolf, P. (2019). Effects of proactive decision making on life satisfaction. *European Journal of Operational Research. 280*(3), 1171-1187.

[29] Silver, N. (2012). The signal and the noise: Why so many predictions fail– but some don't. New York, NY: Penguin Press.

[30] Survival of the Best Fit (SOBF) (2019). Retrieved from https://www.survivalofthebestfit.com/.

[31] Schwaninger, I., Fitzpatrick, G., & Weiss, A. (2019). Exploring Trust in Human-Agent Collaboration. *Proceedings of the 17th European Conference on Computer-Supported Cooperative Work: The International Venue on Practice-centred Computing and the Design of Cooperation Technologies – Exploratory Papers, Reports of the European Society for Socially Embedded Technologies* (pp. 2510 – 2591).

[32] van Belkom, R. (2017). In innovation we trust: right? *Masterthesis Brand Design and Reputation Management, EURIB.* 1-91. <unpublished>

[33] van Belkom, R. (2019). Duikboten zwemmen niet. Retrieved from https://detoekomstvanai.nl.

[34] van de Weijer, B. (2018, June 18). Wiskundige Cathy O'Neil waarschuwt voor algoritmen: 'Rechten van individu worden niet beschermd'. Retrieved from https://www.volkskrant.nl/wetenschap/wiskundige-cathy-o-neil-waarschuwt-voor-algoritmen-rechten-van-individu-worden-niet-beschermd~b97a9302/.

[35] Venkatesh, V., Morris, M. G., Davis, G.B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View, *MIS Quarterly. 27*(3).

[36] Wakefield, J. (2018, June 21). The man who was fired by a machine. Retrieved from https://www.bbc.com/news/technology-44561838.

APPENDIX

A. Descriptives of the demographics

Table A1. *Descriptive statistics on the demographic data of participants of The Dismiss Decision*

		Frequency	Percent
Gender	Male	55	24.9
	Female	164	74.2
	Prefer not to say	2	0.9
Age	18-24	111	50.2
	25-35	91	41.2
	36-45	9	4.1
	46-55	5	2.3
	56-65	3	1.4
	Unknown	2	0.9
Educational level	High school	11	5.0
	Vocational education	1	0.5
	Higher vocational education	22	10.0
	University Bachelor's degree	75	33.9
	University Master's degree	96	43.4
	PhD	12	5.4
	Other	4	1.8
Employment status	Student	137	62.0
	Full-time	63	28.5
	Part-time	13	5.9
	Unemployed	5	2.3
	Other	3	1.4
Region	African	3	1.4
	Asian	7	3.2
	Middle East	2	0.9
	European/Western	207	93.7
	Unknown	2	0.9
Total participants		221	100

B. Explorative questions

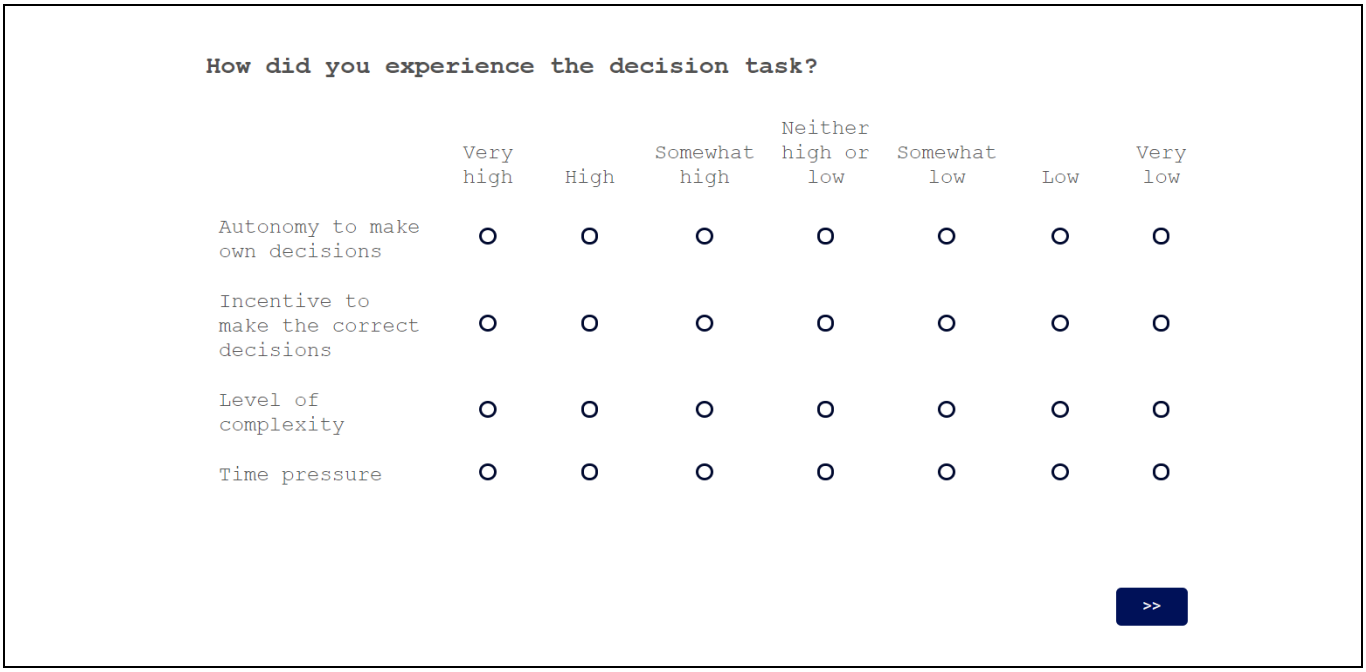

Figure B1. Explorative questions on the perceived experience of the decision task.

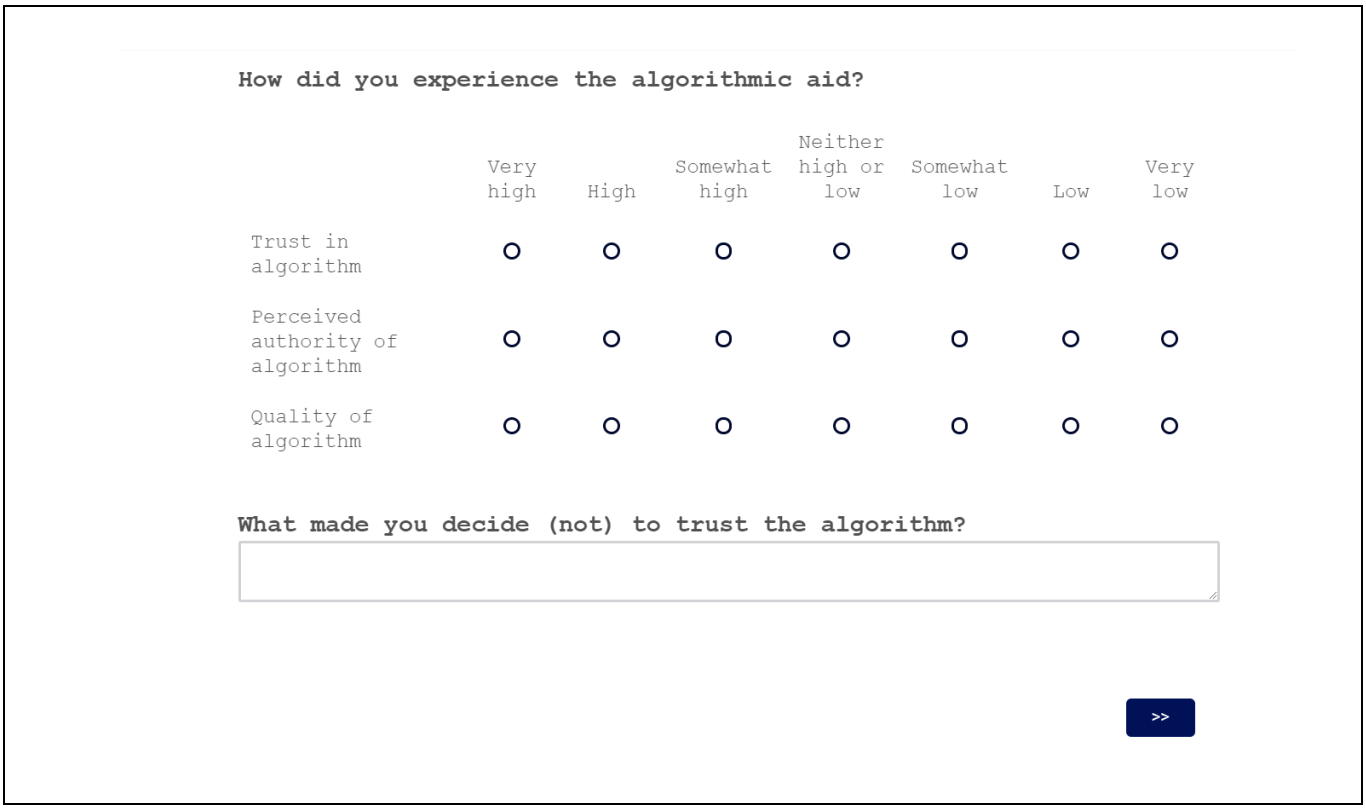

Figure B2. Explorative questions on the experience of the algorithmic aid.

C. Data analysis

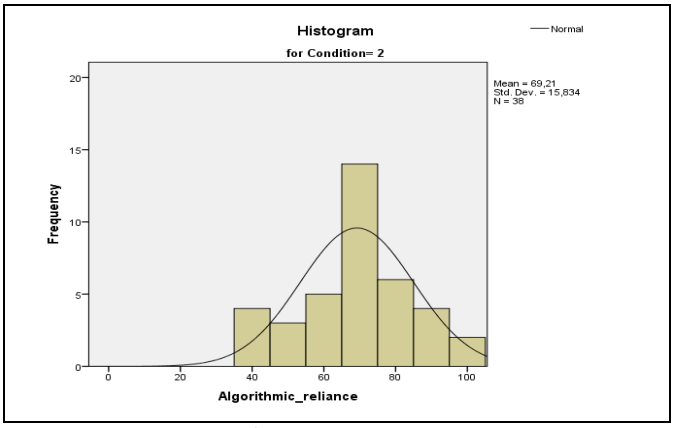

Figure C1. Histogram of algorithmic reliance within condition 2 (control).

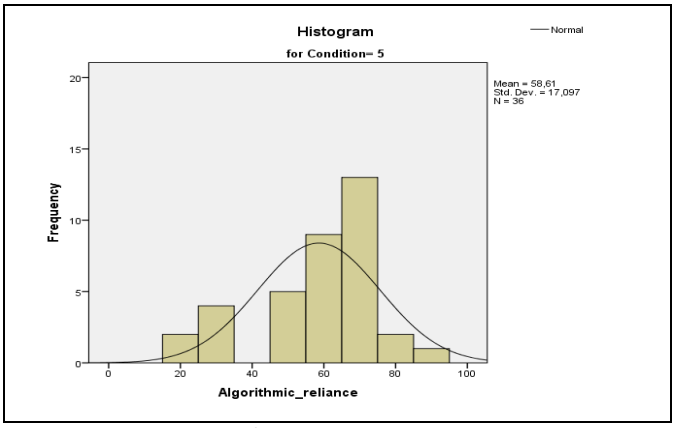

Figure C4. Histogram of algorithmic reliance within condition 5 (social reference).

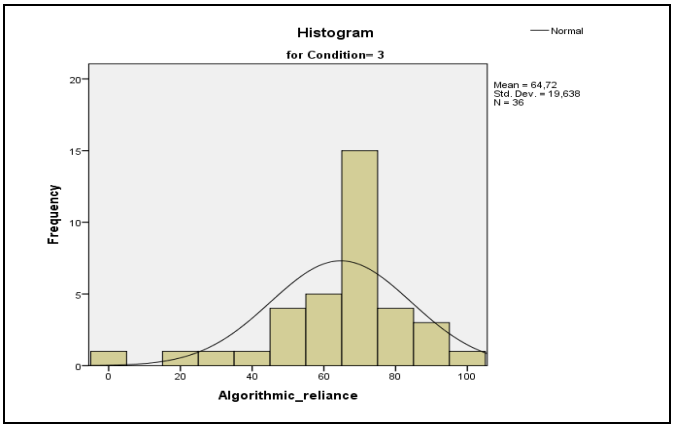

Figure C2. Histogram of algorithmic reliance within condition 3 (self-confidence).

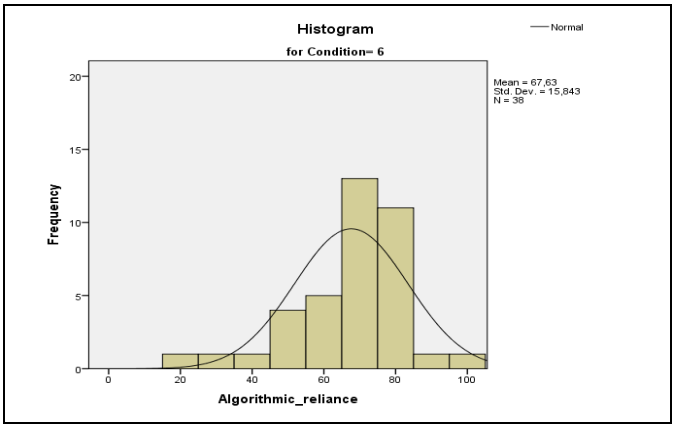

Figure C5. Histogram of algorithmic reliance within condition 6 (qualification of developers).

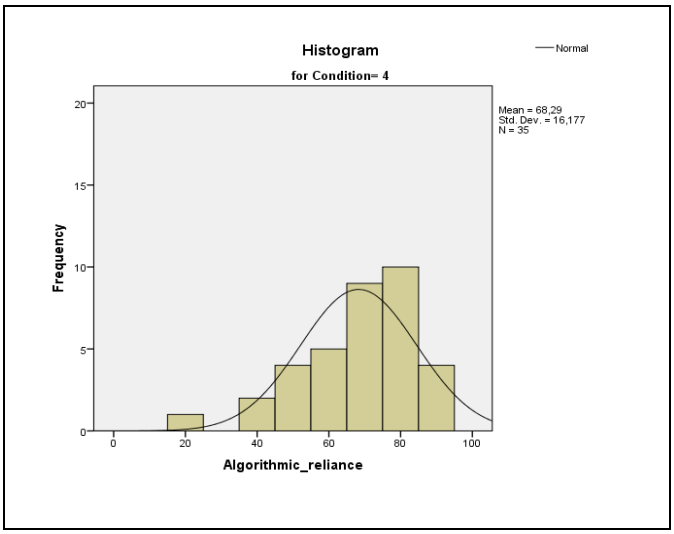

Figure C3. Histogram of algorithmic reliance within condition 4 (historical usage).

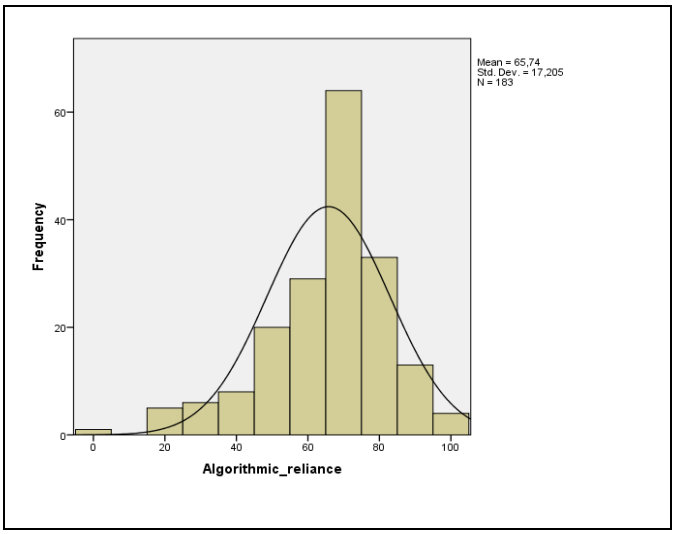

Figure C6. Histogram of algorithmic reliance of all participants within condition 2-6.

Table C1. *Amount of correct decisions per task during baseline control condition (Cb0).*

Phase	Time	Task	Correct decisions	Average correct/phase
1	15 sec	1	94.7%	96.2%
		2	97.4%	
2	12 sec	3	78.9%	74.33%
		4	68.4%	
		5	76.3%	
		6	73.7%	
3	10 sec	7	57.9%	69.73%
		8	60.5%	
		9	68.4%	
		10	65.8%	
		11	92.1%	
		12	73.7%	

Table C2. *Results of independent samples t-tests on the answers of the task-related questions between condition Cb0 (no aid) and C0 (algorithmic aid).*

	Condition	Mean	SD	Sign. (*p*)
Autonomy	Cb0	4.89	1.64	**.076**
	C0	4.24	1.55	
Incentive	Cb0	4.71	1.64	.676
	C0	4.55	1.64	
Complexity	Cb0	4.21	1.55	.224
	C0	4.61	1.24	
Time pressure	Cb0	5.82	1.21	.123
	C0	6.21	0.99	

Table C3. *Results one-way ANOVA of overall task score and answers to explorative questions between conditions C0, C1, C2, C3 and C4*

	Variable	df (between,within)	F	Sign. (*p*)
Overall	Task score	4, 172	1.583	.181
Task related	Autonomy	4, 178	.791	.532
	Incentive	4, 178	1.011	.403
	Complexity	4, 178	3.374	**.011***
	Time pressure	4, 178	1.383	.242
Algorithm related	Trust	4, 178	.778	.541
	Authority	4, 178	1.276	.281
	Quality	4, 178	2.115	**.081**

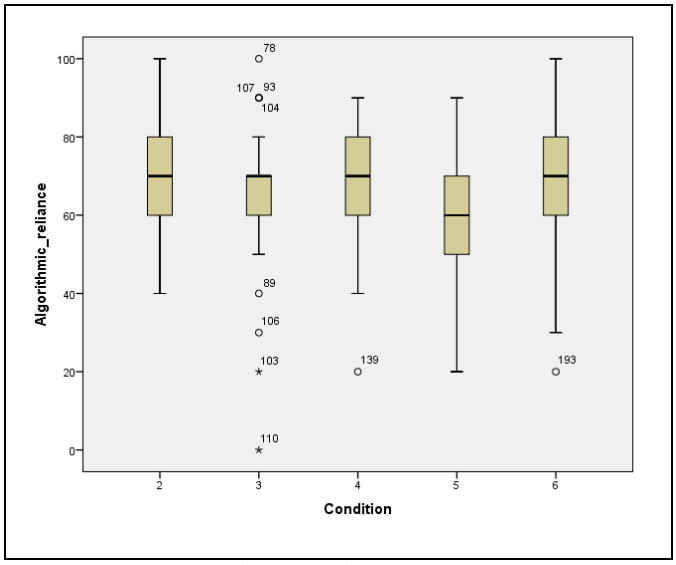

Figure C7. Boxplots of amount of algorithmic reliance per condition.

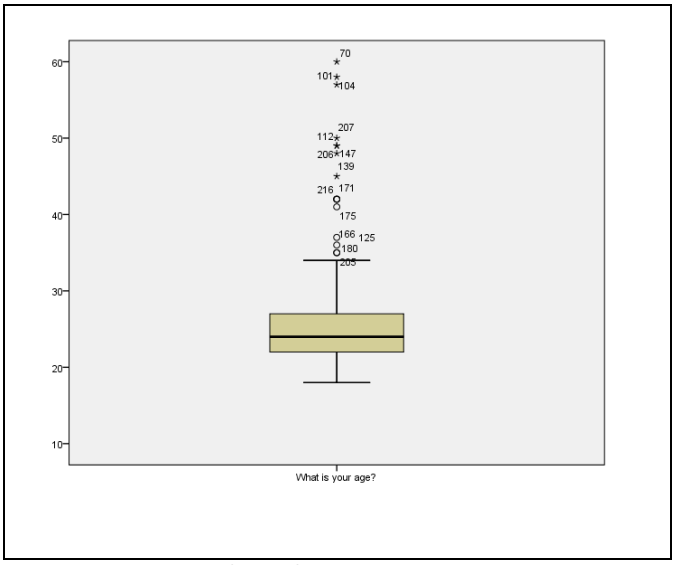

Figure C9. Boxplot of age of the participants.

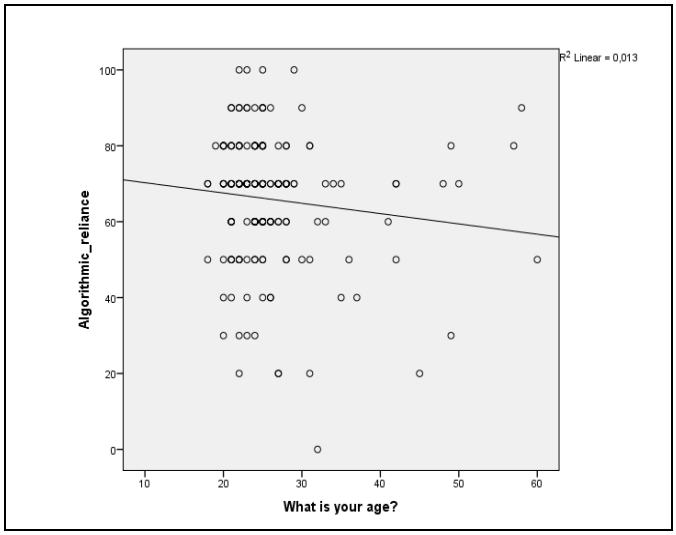

Figure C8. Scatterplot of the relation between age and algorithmic reliance.