Attitudes to decision-making under risk supported by artificial intelligence and humans

Perceived risk, reliability and acceptance

Per Nordahl
2019
Preface

I would like to thank my thesis supervisor, Fredrik Bökman, for his guidance, patience and valuable comments. I would like to thank the participants of the online survey for taking the time to answer the questionnaire and for sharing their thoughts on artificial intelligence (AI) decision-making.
Abstract

The purpose of this investigation was to explore how decision situations with varying degrees of perceived risk affect people’s attitudes to human and artificial intelligence (AI) decision-making support. While previous studies have focused on the trust, fairness, reliability and fear of artificial intelligence, robots and algorithms in relation to decision support, the risk inherent in the decision situation has been largely ignored. An online survey with a mixed approach was conducted to investigate artificial intelligence and human decision support in risky situations. Two scenarios were presented to the survey participants. In the scenario where the perceived situational risk was low, selecting a restaurant, people expressed a positive attitude towards relying on and accepting recommendations provided by an AI. In contrast, in the perceived high-risk scenario, purchasing a home, people expressed an equal reluctance to rely on or accept both AI and human recommendations. The limitations of this investigation are primarily related to the challenges of creating a common understanding of concepts such as AI and a relatively homogenous survey group. The implication of this study is that AI may currently be best applied to situations characterized by perceived low risk if the intention is to convince people to rely on and accept AI recommendations, and in the future if AI becomes autonomous, to accept decisions.
# Table of Contents

1  Introduction ............................................................................................................. 1  
1.1  Questions to study .............................................................................................. 3  
1.2  Overview of thesis ............................................................................................... 4  
2  Literature review .................................................................................................... 5  
2.1  What is artificial intelligence? ............................................................................. 5  
2.2  Decision-making and risk ................................................................................... 6  
2.2.1  Decision-making ............................................................................................. 6  
2.2.2  Decision-making with artificial intelligence .................................................... 7  
2.2.3  Risk and perceived risk .................................................................................. 8  
2.3  Perspectives on decision-making and artificial intelligence .............................. 9  
2.3.1  Algorithm aversion ......................................................................................... 10  
2.3.2  Trust in and fairness of algorithms ................................................................. 11  
2.3.3  Trust in decision support systems .................................................................. 12  
2.3.4  Robots and misplaced trust .......................................................................... 13  
2.3.5  AI support for decision-making incapacity ...................................................... 15  
2.3.6  Fear of robots and AI ..................................................................................... 16  
2.3.7  Negative attitudes towards robots .................................................................. 17  
2.3.8  AI and job elimination ................................................................................... 17  
2.3.9  Comparative advantage, AI vs. human ........................................................... 18  
2.3.10  Risk perception and autonomous vehicles ..................................................... 20  
2.3.11  Ethical AI ...................................................................................................... 21  
2.3.12  Prospect theory and human “irrationality” .................................................... 23  
2.4  Summary and relation to hypotheses .................................................................. 24  
3  Design of study ...................................................................................................... 25  
3.1  Method ................................................................................................................ 25  
3.1.1  Chosen risky scenarios .................................................................................... 25  
3.1.2  Data collection and survey structure ............................................................... 26  
3.1.3  Sampling ......................................................................................................... 28  
3.1.4  Data analysis ................................................................................................... 28  
3.2  Research ethics ................................................................................................... 28  
4  Survey results ....................................................................................................... 30  
4.1  Survey sampling and response ......................................................................... 30  
4.2  Participant coding ............................................................................................... 30  
4.3  Attitudes to risk .................................................................................................. 32  
4.4  Attitudes to decision support ............................................................................. 33  
4.5  Home purchasing scenario results .................................................................... 34  
4.6  Restaurant selection scenario results .................................................................. 37  
4.7  Qualitative data .................................................................................................. 40  
5  Analysis ................................................................................................................ 42
5.1 Participant overall risk evaluation of scenarios ............................................................ 42
5.2 Risk aversion, reliability and acceptance ................................................................. 42
5.3 Negative attitudes to AI, reliability and acceptance ................................................... 43
5.4 Perceived risk and AI reliability ................................................................................ 43
5.5 Impact and AI reliability ............................................................................................ 44
5.6 Acceptance of AI recommendations .......................................................................... 45
5.7 Gender and age vs. reliability and acceptance ............................................................ 46
5.8 Within scenario reliability and acceptance ................................................................. 47
5.9 Between scenario reliability and acceptance ............................................................... 48
6  Discussion ..................................................................................................................... 50
6.1 Risk aversion and negative attitudes to AI ................................................................. 50
6.2 Perceived risk and AI reliability ................................................................................ 51
6.3 Acceptance of AI recommendations .......................................................................... 52
6.4 Situational risk, reliability and acceptance ............................................................... 52
6.5 Gender and age perspectives ..................................................................................... 53
6.6 Perspectives on chosen method ................................................................................ 54
7  Conclusion ................................................................................................................... 56
7.1 Study outcomes ......................................................................................................... 56
7.2 Future research ......................................................................................................... 57
References ..................................................................................................................... 58
Appendix A .................................................................................................................. A1
1 Introduction

The purpose of this chapter is to introduce the topic that was studied, describe the problem and to present avenues of investigation. A cursory overview of the field is presented and the aim, research question and hypotheses are formulated.

"Science fiction then is the fiction of revolutions. Revolutions in time, space, medicine, travel, and thought... Above all, science fiction is the fiction of warm-blooded human men and women sometimes elevated and sometimes crushed by their machines." (Ray Bradbury, 1974 via Sargeant, 2018)

Artificial Intelligence (AI) is currently a hot topic from political, economic, socio-cultural, technological, environmental and legal perspectives. AI is expected to have a huge impact on human endeavors going forward, but the ways in which AI will be implemented, not only technologically but also culturally are highly uncertain.

What is Artificial Intelligence? According to Harmon (2018) AI is simply an evolution of existing software and hardware paradigms that have existed since the 1970's, such as reasoning systems, vision- and language processing. New programming techniques are allowing computers to do new things, but the main breakthrough is the availability of sufficient processing power to realize techniques that were previously impractical. Harmon (2018) describes three types of human tasks; physical work, mental work and interpersonal work. His opinion is that the first type is the easiest to replace with artificial intelligence while the latter may be the most difficult.

Agrawal et al. (2016) argue that improvements to machine intelligence will decrease the value of human prediction skills, as machines are potentially better and cheaper in this area, but that in turn the value of human judgement skills will increase. This argument assumes that judgement and prediction are (economic) complements to each other and when the cost of prediction decreases the demand for judgement increases. While this is an elegant argument it is far from certain. They (Agrawal et al., 2017) additionally state that as prediction becomes cheaper, demand for decision-making will increase resulting in more total predictions being generated. Therefore, there will be more opportunities for humans to exercise judgement.

How will humans react to this increased need for judgement? Will we choose to seize those opportunities (Broughham and Haar, 2018:241) as they become available or will we instead succumb to judgement overload? To this end Simonite (2017) infers that

"making AI systems that can soak up goals and motivations from humans has emerged as a major theme in the expanding project of making machines that are both safe and smart."
So where are we now? Claudé & Combe (2018:3) found that (even partially) autonomous AI is not very common at the moment; rather it supports and augments human beings in decision-making, but who knows what the future will hold?

According to Brynjolfsson & McAfee (2014:15, 28-29) when humans attempt to compete with digital labor they should focus on those areas where they have a comparative advantage over computers, letting computers do what they are suited for best. For example, they highlight Moravec's paradox, that low-level sensorimotor skills require huge amounts of computational resources while high-level reasoning requires very little computation. This is summarized by AI researcher Steven Pinker as "the hard problems are easy and the easy problems are hard." For humans this requires us to reflect, are questions of judgement easy problems or hard? Further, do universal AI truths (solutions) exist or are they different and specific depending on a certain time and place (or problem) (Baggini, 2018)?

My generation of teens preferred real-life friends. It seems that this generation of teens prefers online friends (Pew Research Center via RT.com, 2015). How likely is that some future generation of teens will prefer artificial friends and what will that mean for human to human relationships and the future of decision-making?

Will humans trust AIs to decide for them? Will humans accept these decisions? AIs make mistakes, as do humans. Are AIs reliable in their decision-making? If an AI makes a mistake will humans be willing to forgive and forget or will they hold a grudge? Will they blame the humans behind the AI? Lindström (2017) reports that in fact we may prefer for AIs to be flawed and make (some) mistakes because it makes us recognize our own imperfections and be more willing to accept interactions and cooperation with AI. But what if it’s the human making the mistakes? Dietvorst et al. (2014:1) found that

"people are especially averse to algorithmic forecasters after seeing them perform, even when they see them outperform a human forecaster."

Would irrational human’s interests therefore be better served by acquiescing to a rational AI?

According to Tegmark (2017:33-34, 37) the conversation about AI is an important one, both in terms of urgency and impact. Some AI researchers worry that focusing on the risks of AI is a distraction that could slow down progress, while at the same time mainstream AI researchers have sometimes misunderstood the concept of AI-safety research.

Clearly attitudes to artificial intelligence decision-making support are an area that warrants further study (Greene, 2018). This investigation contributes by deepening
our understanding of the varying attitudes to AI decision-making support in relation to perceived risk.

### 1.1 Questions to study

This study was grounded in decision theory focusing on perspectives related to varying degrees of risk in the decision situation. This relates to studies on autonomous driving based on risk (Brell et al., 2018) and research into decision aids (Madhavan and Wiegmann, 2007) and human and AI collaboration (Jarrahi, 2018) grounded in decision theory.

This contrasts to studies which looked at social engineering (Aroyo et al., 2018), algorithmic aversion (Dietvorst et al., 2014), managerial decisions (Lee, 2018) and negative attitudes in human-robot interactions (Nomura et al., 2006) from a psychological perspective. Additionally, this investigation has touch points with ethical considerations such as medical decision support (Lamanna and Lauren, 2018) and autonomous decision-making (Etzioni and Etzioni, 2017) as well as the fear of robots and AI (Liang and Lee, 2017) based in sociology and unemployment caused by AI (Broughham and Haar, 2018) from an organizational perspective. As Claudé & Combe (2018:3) have found, AI is currently mainly used to support and augment human beings in decision-making and isn't yet commonly autonomous (as of this writing).

The aim of the study was to explore decision situations with varying degrees of perceived risk related to human and machine decision maker support. It is normal for researchers to incentivize participants by paying them to participate and sometimes rewarding them for their performance (Dietvorst et al., 2014 and Lee, 2018). It is uncommon that participants have a personal stake in the decision situations themselves, with some exceptions (Aroyo et al., 2018). The lack of personal risk in the decision situation and how it impacts the perceived reliability of and trust in AI is a problem worth exploring further. To this end this study refrained from providing paid incentives to participants and only created risky situations, in the form of thought experiments, for the participants to explore.

The question this paper attempts to answer is the following: How are people’s attitudes to artificial intelligence vs. human decision-making support affected by the level of perceived risk in the decision situation?

Several hypotheses were formulated to support answering this question. The literature review and previous research was used to guide and inspire the creation of the following statements:

**H1** Risk averse participants will be less likely to rely on or accept an AI recommendation.
H2 Participants with a negative attitude to AI will be less likely to rely on or accept an AI recommendation

H3A Participants will be more likely to rely on human rather than AI support in a decision situation characterized by high perceived risk

H3B Participants will be equally likely to rely on human and AI support in a decision situation characterized by low perceived risk

H4A Participants will be more likely to rely on human rather than AI support in a decision situation characterized by high impact

H4B Participants will be equally likely to rely on human and AI support in a decision situation characterized by low impact

H5 Participants will overall be more likely to accept a human than an AI recommendation

H6A Women will be less likely to rely on or accept an AI recommendation

H6B Older age groups (55+) will be less likely to rely on or accept an AI recommendation

1.2 Overview of thesis

The first chapter introduces the topic that was studied, describes the problem and presents the avenues of investigation. An overview of AI and decision-making is presented and the aim, research question and hypotheses are formulated. Chapter two explores current theory and earlier research in the context of human attitudes to artificial intelligence decision-making support under risk. The third chapter describes the design of the study and the implementation of the chosen method. Chapter four presents the results of the online survey. The fifth chapter analyzes the data using statistical tools. Chapter six discusses the findings presented in the analysis chapter. The hypotheses are revisited, and the implications of the findings are presented as well as perspectives on the chosen method. The final chapter presents an overall answer to the investigated question and the tested hypotheses. The contributions to theory are described including suggestions for future research. The paper concludes with a list of references and an appendix containing the survey that was published and made available to the participants in the online tool.
2 Literature review

The purpose of this chapter is to explore current theory and earlier research in the context of human attitudes to artificial intelligence decision-making support under risk.

2.1 What is artificial intelligence?

There are many different views and definitions of artificial intelligence. Further, the different definitions used by researchers may not be particularly well aligned with the concept of AI in the public conscious.

According to Russell and Norvig (2010:1) the term artificial intelligence (AI) was coined in 1956. The authors provide 4 different definitions of AI. They differentiate between thinking vs. acting definitions as well human vs. rational definitions. In Table 1 below these definitions are explained:

Table 1 - Definitions of AI, adapted from Russell and Norvig (2010:2)

<table>
<thead>
<tr>
<th>Thinking Human</th>
<th>Thinking Rational</th>
</tr>
</thead>
<tbody>
<tr>
<td>“[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . .” (Bellman, 1978)</td>
<td>“The study of the computations that make it possible to perceive, reason, and act.” (Winston, 1992)</td>
</tr>
<tr>
<td>Acting Human</td>
<td>Acting Rational</td>
</tr>
<tr>
<td>“The art of creating machines that perform functions that require intelligence when performed by people.” (Kurzweil, 1990)</td>
<td>“Computational Intelligence is the study of the design of intelligent agents.” (Poole et al., 1998)</td>
</tr>
</tbody>
</table>

The acting human approach uses the Turing Test (devised by Alan Turing in 1950) as the starting point for defining AI. To pass a Turing Test a computer would need to fool a human interrogator into thinking he or she was communicating with another person. To achieve this, the computer would need to possess certain capabilities such as natural language processing, knowledge representation, automated reasoning and machine learning. In addition, it might need computer vision (and other sensors) and robotics to be capable of perceiving and manipulating the physical world around it (Russell and Norvig, 2010:2-3).

The thinking human approach in defining AI approaches the subject from the perspective of cognitive science, combining computer modeling with theories from psychology. To create a computer that thinks like a human it must be possible to determine how humans think (Russell and Norvig, 2010:3).
The thinking rational definitions of AI (Russell and Norvig, 2010:4) are grounded in the study of logic. The approach uses logic notation, as opposed to arithmetic notation, to attempt to solve problems. The challenges with using this approach include limitations when handling uncertainty and the applicability of theoretical solutions to practical, real-world problems.

The final group of definitions is the acting rational approach which utilizes the perspective of the rational agent. A rational agent acts so that the best outcome is achieved, or when subjected to uncertainty, the best expected outcome. This approach adopts the skills needed to fulfill the Turing Test and additionally the authors propose it to be advantageous because it provides multiple avenues for success and better lends itself to scientific development compared to approaches from the human dimension (Russell and Norvig, 2010:4-5).

The authors (Russell and Norvig, 2010:9-10) note that decision theory, as an offshoot of economics, provides some of the philosophical underpinnings for AI. They define decision theory as combining probability theory and utility theory into a framework for making decisions under uncertainty. They note that decision theory is appropriate for decisions in which agents’ actions are independent of each other while game theory is suitable when one player's actions can affect the utility of another. While decision theory provides unambiguous prescriptions for selecting actions, game theory instead posits that sometimes the best course of action for an agent may be random (or appear to be). Further, operation research addresses the question of delayed payoffs, when the consequences of a decision are not immediate or if the result is generated by a sequence of events. Further work in this area developed models for making decisions that are good enough (satisficing) rather than perfect (optimal). This has been helpful since it seems to more accurately reflect actual human behavior.

### 2.2 Decision-making and risk

According to the literature review conducted by Mardani et al. (2015:548-550) on decision-making techniques and their applications many different tools for modeling and aiding decision-making exist today. The authors conclude that choosing a suitable approach is based on the people involved, the desired outcomes as well as the availability of time and information. They identify the ability to address the problems of conflicting interests as the most important advantage of decision-making methods.

#### 2.2.1 Decision-making

Why are decisions hard? Clemen and Reilly (2014:3) describe several factors that contribute to complicating decision-making including uncertainty, complexity,
competing viewpoints and multiple objectives. While good decisions can have unlucky outcomes, decision-making can be improved by carefully understanding and considering the different issues related to a problem (Clemen and Reilly, 2014:5). Decision analysis can be used to structure a problem to identify objectives, alternative courses of action and trade-offs. After a decision has been made, decision analysis can be used to justify a chosen alternative (Clemen and Reilly, 2014:7).

Hammond et al (1999:5) describe a systematic decision-making process called the PrOACT-method. The process consists of 5 core elements (Figure 1 below) and 3 additional elements. The core elements are problem, objectives, alternatives, consequences and trade-offs. The additional elements are uncertainty, risk tolerance and linked decisions. According to the authors the method is applicable to all types of decisions.

![Figure 1 - PrOACT model, adapted from Hammond et al. (1999:5)](image)

The first additional element of the PrOACT-method is uncertainty. Uncertainty means that the decision maker doesn't fully know or understand the consequences of a decision until after the decision has been made. The authors explain that an uncertain decision should be judged and evaluated based on the quality of the decision-making, not on the quality of the actual consequences (Hammond et al, 1999:109–111). Risk tolerance is the second additional element of the PrOACT-method. It is important since people have varying degrees of willingness to take risk to attain better consequences (Hammond et al, 1999:137). The final additional element of the PrOACT-method is linked decisions. Sometimes an alternative chosen today creates alternatives that become available tomorrow and can influence how attractive future alternatives are (Hammond et al, 1999:163).

### 2.2.2 Decision-making with artificial intelligence

How would one go about designing an AI to make simple and complex decisions? According to Russell and Norvig (2010:636) the construction of an AI agent capable of selecting actions to maximize performance would be based on combining probability theory, utility theory and decision theory. Probability theory would provide an evidence-based description of what the AI should believe. Utility theory would describe the wants (preferences) of the agent through a utility function and decision theory would combine the two previous theories to provide a description of what the agent should do. Decision networks (aka influence diagrams) could be used to formalize the descriptions of and solutions to the decision problems. To design a ra-
tional agent, decision theory would be needed to construct a system in which the decisions made would consider all possible alternatives. The AI agent, utilizing perfect information, would have the capability of selecting the alternative that results in the best expected outcome. Since it is highly unlikely that an AI would have access to all the (available, relevant) information necessary to make perfectly rational decisions the value of information would have to be considered (Russell and Norvig, 2010:628-629). This would support the AI agent in selecting which information to acquire and

"is defined as the expected improvement in utility compared with making a decision without the information." (Russell and Norvig, 2010:636)

To make more complex decisions, for example in stochastic environments of sequential decision problems, further capabilities would be required of the AI agent such as Markov decision processes (MDP) and dynamic decision networks (Russell and Norvig, 2010:645, 684-685). If multiple agents were competing to maximize their utility, game theory would need to be incorporated into the design of the system (Russell and Norvig, 2010:666).

2.2.3 Risk and perceived risk

Risk is normally associated with some type of decision and analyzed based on what might happen, for example what is the likelihood that an event will occur and what is the valuation of the impact of the risk’s consequences?

According to Sjöberg and Thedéen (2003:16-17) there are four common definitions of risk: the probability of a harmful event occurring, the product of the probability of a harmful event and its impact (i.e. number of people killed in an accident), the variation in outcome if a certain action is taken and finally, perceived risk. Perceived risk is defined as how large or small a person believes a risk to be, with the risk itself defined by what the person believes the word to mean to them.

According to Clemen and Reilly (2014:8) for some people risk means the chance of monetary loss while for others it refers to situations in which health or the environment may be potentially damaged.

When evaluating risk, the actual risk or objective size of a risk is what really matters, but humans are subjective and tend to act and react individually. In the real world people tend to rely on perceived risk and they tend to underestimate large risks and overestimate small ones. Additionally, the public tends to believe that risks are greater than the experts do. On the other hand, the experts are not always correct in their risk judgements (Drottz-Sjöberg and Sjöberg, 2003:316).

According to Slovic and Peters (2006:322) humans primarily perceive and act on risk in two ways. People react to danger intuitively and instinctively, which can be
referred to risk as feelings. In contrast, risk as thought, in the form of analysis, utilizes reason and logic in order to manage risk. In practice people often apply intuition and experimentation when managing risk. They tend to rely on an affect heuristic, as this is normally easier and quicker and possibly a more efficient way to relate to an uncertain and complex world.

Earlier research (Slovic and Peters, 2006:323) indicates that:

"people judge a risk not only by what they think about it but also by how they feel about it. If their feelings toward an activity are favorable, they tend to judge the risks as low and the benefits as high; if their feelings toward the activity are unfavorable, they tend to make the opposite judgment."

According to the psychometric model (Drottz-Sjöberg and Sjöberg, 2003:319) three factors can be identified that are necessary for people to make meaningful risk judgements: number of victims, if the risk is new or not and fear. Other perspectives that may need to be considered include risk availability (Drottz-Sjöberg and Sjöberg, 2003:317), free will (Drottz-Sjöberg and Sjöberg, 2003:318) as well as if the risk is personal or general (Drottz-Sjöberg and Sjöberg, 2003:325). The negative consequences can be expressed as lost production, willingness to pay (to avoid a risk) and compensation claims which all may need to be considered differently (Mattsson, 2003:359).

An additional perspective is risk communication where credibility and stakeholder perspective are important factors (Drottz-Sjöberg and Sjöberg, 2003:334). In the continuous process of risk communication, it’s important to use the target groups possibilities and limitations as a starting point for a successful discussion of risk. Since probability is often perceived as a difficult and theoretical concept that few people fully understand (Sjöberg and Thedéen, 2003:16-17), risk definitions containing probability should be generally avoided when communicating with the public (also see Clemen and Reilly, 2014:711).

2.3 Perspectives on decision-making and artificial intelligence

According to Ariely (2011:9-10), by getting a better understanding of the different irrational forces that shape our decision-making, we as humans can possibly overcome our biases and make better decisions. To this end we can

"redesign our working and living environments in ways that are naturally more compatible with what we can and cannot do."

Potentially AI can contribute to this redesign of our decision environment and enable us as humans to do things we naturally couldn't.
2.3.1 Algorithm aversion

In their studies Dietvorst et al. (2014:1) identified a concept they dubbed “algorithm aversion”. The authors define this phenomenon as occurring in a forecasting situation, when algorithms show superior performance compared to humans, the rational thing for a human decision maker would be to select the forecast generated by the algorithm. But this is often not the case and humans irrationally select the human forecast, since they show aversion to the one generated by the algorithm. The term "algorithm" is used by the authors to refer to forecasting procedures that include statistical models, decision rules and other mechanical processes, which seems to indicate that it also covers AI, although the term AI is not specifically used.

Dietvorst et al. (2014:2) conducted five studies that showed that seeing an algorithm make errors would make people less likely to choose the algorithm compared to a human forecaster. This was found to occur even if the algorithm was shown to outperform a human forecaster and independent of if the human forecasts were generated by the participants themselves or another anonymous participant. In three of the five studies (Dietvorst et al., 2014:2) participants were asked to predict how well MBA students had performed in an MBA program based on admissions data. The participants where further asked to select a human judge or a statistical model to support their prediction. Before being allowed to choose which decision support to use the participants were shown the performance of the human, the performance of the algorithm, neither or both. Participants were incentivized with a fee for showing up and could earn additional money based on their forecast performance. The incentives were the same for all three studies and participants. In the remaining two studies (Dietvorst et al., 2014:5) participants were asked to predict the rank of US states based on the number of departing airline passengers. Participants in these studies were incentivized with $1 for completing the study and $1 for forecasting performance (Dietvorst et al., 2014:3). In all five studies “the machine” outperformed the human forecasters.

One interesting result from the studies is the impact on participant’s confidence ratings in the prediction ability of humans vs. machines. It indicated that participants learned more from errors made by the statistical model than those made by humans. When humans made relatively large mistakes, this did not impact the participant’s confidence in the human forecasts. On the other hand, seeing the model make relatively minor mistakes had a consistently negative impact on the human confidence in the model’s predictions (Dietvorst et al., 2014:8). Besides confidence the authors additionally measured beliefs about the human's and model's predictive ability. The overall result was that if participants saw the model perform, they were less optimistic about it, but there were some additional interesting details. Participants thought the model would be better than humans at avoiding obvious errors and ap-
propriately and consistently weighing attributes and information. Regarding belief in human’s predictive ability the participants thought humans would be better than the model regarding improving with practice, learning from mistakes and finding novel solutions (Dietvorst et al., 2014:9-10). In the decision situations that the authors explored there was no personal stake in the decision situations for the participants themselves. The researchers instead incentivized the participants by paying them to participate and rewarded them for forecast accuracy.

2.3.2 Trust in and fairness of algorithms

According Lee (2018:1) we currently don’t fully comprehend how people perceive decision made by humans compared to decisions made by algorithms. To this end Lee (2018:2) explored how people feel about managerial decisions being taken over by algorithms and if the decisions made by algorithms were more trustworthy and fairer or less trustworthy and fair compared to those made by humans, independent of the decision outcome. The author formulated three hypotheses: that algorithms are perceived as fairer in decisions that need human skills but not those that need mechanical skills, that algorithms are trusted equally as humans in decisions that need mechanical skills but less trusted in decisions that need human skills and that decisions by an algorithm cause a less emotional response than those made by a human (Lee, 2018:4-5). An online experiment was conducted utilizing four different third person managerial decision scenarios requiring either mechanical (e.g. data processing) or human (e.g. judgement, emotion) skills and where the described decision maker was either an algorithm or a human. The participants were paid for taking part in the experiment.

On the topic of fairness, only partial support for the hypothesis was found by Lee (2018:7-9). For mechanical tasks the results were that both human and algorithmic decision makers were equally fair according to participants, but on human tasks the human decision maker was judged to be fairer. The human fairness was judged by participants based on managerial position, qualifications and authority while algorithmic fairness was based on a perception of the algorithm applying a consistent set of rules resulting in a lack of bias. The algorithm was deemed less fair in circumstances where it was judged not to consider human intuition, contexts and concerns, not explicitly apparent in the decision situation.

Regarding trust it was found (Lee, 2018:9-10) that the hypothesis was supported. Human and algorithmic decisions were equally trusted for mechanical skills. The participant’s judgements of trust were like those found for fairness, but they additionally highlighted the algorithms potential for bias while ignoring the human’s potential for mistakes. Humans were found to be more trustworthy regarding decisions requiring human skills. The reasons were similar as with fairness with participants citing the managerial position and therefore implicit skill of the human com-
pared to the assumed poor design and lack of exception handling (human circumstances) for the algorithm.

When reviewing emotional response (Lee, 2018:10-11), the hypothesis was refuted. For decisions relating to mechanical skills the emotional response was the same but for human skill decisions the participants were more negative towards the algorithm. The human decision-maker was judged by participants as being able to provide social recognition while algorithms were perceived to remove human agency and cause a feeling of being watched. Some thought that being evaluated by a machine was disrespectful and demeaning, related to a lack of trust and fairness.

Overall the author concludes that people perceive decisions made by algorithms as

"less fair, less trustworthy, and more likely to evoke negative emotion for tasks that people think require uniquely human skills." (Lee, 2018:14)

As with Dietvorst et al. (2014) in the decision situations that the author explored there was no personal stake in the decision situations for the participants themselves (third person scenarios).

2.3.3 Trust in decision support systems

Madhavan and Wiegmann (2007:773) conducted two experiments to study human perceptions of decision support systems (DSS). As automation becomes more common the authors describe the role of the human operator as changing from primary controller into an active teammate that shares control responsibility with the automation. Decision support systems have a history going back to the 1950's and have, as the authors describe, been designed to behave and interact in ways that are like people (Thinking & Acting Human, Russell and Norvig, 2010:1) and performing functions that are difficult or impossible for independent humans.

Earlier research suggested the importance of trust in the decision-making processes of human-machine teams. Some people tend to have a higher initial trust of automation compared to humans due to a "perfect automation schema" bias, therefore with a higher sensitivity towards automation errors, resulting in reduced trust and confidence. The researchers criticize earlier studies for not providing participants with sufficient background context regarding the level of expertise of their human and machine advisors. Additional weaknesses with earlier research include not sufficiently allowing the machine to prove its "worth" (reliability) to the human before being evaluated, which is a foundation for building trust (Madhavan and Wiegmann, 2007:774).

The purpose of the first study (Madhavan and Wiegmann, 2007:775-776) was to evaluate participants preconceived notions of automated vs. human decision support, together with differences in levels of expertise, before performing an actual
task together. Participants were asked to rate their level of trust and the perceived reliability of four different advisors. The authors predicted that the level of expertise of the decision support would influence the participants' trust. For novice level decision-support the automation would be trusted more and for expert level decision-support the human would be perceived as more trustworthy. Additionally, they thought automated advisors would be viewed as more reliable regardless of expertise. These hypotheses were supported by the results of the study.

The second study (Madhavan and Wiegmann, 2007:777-778) looked at a DSS or human providing support in an airport security luggage screening scenario. The participants were told that the advice was coming from either an automated system or a human (in truth all advice was coming from an automated system) with varying stated levels of expertise (novice, expert) and reliability (high, low). The researchers expected participants to prefer automated aids to humans when reliability was high and to discount the automation's level of expertise if errors occurred.

These hypotheses were largely confirmed. For expert humans with low reliability the participants seemed willing to place trust in them because of their pedigree (expert), however this didn't translate to the expert automated systems that made mistakes (Madhavan and Wiegmann, 2007:781-782). On the whole novice advisors were trusted less than the experts.

Overall Madhavan and Wiegmann (2007:783) explain that barriers to humans trusting DSS may be influenced "by incorporating ‘humanlike’ characteristics into the design of automated systems", but that this may still not be enough to overcome preconceived notions and biases. The authors admit that there were no specific consequences for wrong decisions (for the participants) in their studies as with Dietvorst et al. (2014) and (Lee, 2018).

2.3.4 Robots and misplaced trust

While other researchers (Dietvorst et al. (2014), Lee (2018), Madhavan and Wiegmann (2007)) have investigated the lack of trust or barriers to trust in AI, Aroyo et al. (2018:3701) instead explore situations where humans potentially misplace trust or place over-trust in AI. Other studies have investigated trust in human-robot interaction (HRI) based on environmental factors, characteristics of robots such as efficiency, performance and reliability as well as transparency (why the AI behaves in a certain way). The authors study how trust in robots could be exploited in social engineering (SE) situations. Over-trust can be caused by an inappropriate reliance on a malfunctioning or non-transparent system, with negative consequences such as a loss in profitability or compromised safety, going so far as humans accepting bribes from a robot.
Social engineering in a human-to-human context is the psychological manipulation of people to persuade them to perform certain actions or divulge protected information. There is a risk that social engineering techniques could be extended to human-robot interactions with the added advantage for the attacker that robots can move around while transmitting audio and video, impacting everything from intelligent toys to robotic surgeons (Aroyo et al., 2018:3701-3702).

The authors (Aroyo et al., 2018:3702) proposed an experiment based on Kevin Mitnick’s social engineering model to evaluate if a robot can collect information from humans, build trust and rapport with them and then exploit that trust to persuade the humans to perform actions for them. The researchers used a humanoid robot called iCub to interview 61 participants. Thereafter iCub and the participants played a game together where the participants could win money. Finally, iCub tried to exploit the participants trust by suggesting the money won be gambled, based on the robot’s advice. Three hypotheses were proposed: people generally not prone to SE or with a negative attitude towards robots would be less likely to share information, all participants would rely on the robot during the game but risk averse people would not accept the robot’s proposal to gamble and people who won the game and successfully gambled would see the biggest improvement in rapport with the robot.

The first hypothesis was rejected by the findings (Aroyo et al., 2018:3704). The study showed that 92% of participants replied to all the questions regardless of their proneness to SE or their attitudes towards robots. In the second phase (Aroyo et al., 2018:3705) only 61% of participants managed to complete the game and could gamble. Of that group 43% successfully won the gamble. The first part of the second hypothesis was partially supported as the participants complied with the robot’s suggestions but only relied on its help when the game became more difficult. The second part of the second hypothesis was rejected. The robot managed to convince even the risk averse people to gamble. The third hypothesis was also rejected (Aroyo et al., 2018:3706). The greatest changes in positive perception of the robot were recorded for participants who lost the gamble indicating that even when unsuccessful in supporting a human, a robot may be capable of building empathy and rapport.

Overall the results of the study revealed that robots can build trust, influence people to share personal information and to influence humans to conform to suggestions (Aroyo et al., 2018:3706). This shows that social engineering techniques can be successfully extended from human-to-human interactions to the human-robot domain, potentially facilitating crimes such as identity theft. A potential limitation of the study is that the laboratory environment may have made participants feel safe, causing them to reveal and trust more than they would have been normally inclined to
do. On the other hand, most social engineering attacks occur in situations where the victim feels safe such as at home or in the office (Aroyo et al., 2018:3707). In contrast to other researchers, Aroyo et al. (2018) did evaluate the risk tolerance of the participants and explored a decision situation where there was a real risk of winning or losing something of some consequence (£7.50 UK).

2.3.5 AI support for decision-making incapacity

Lamanna and Lauren (2018:902) argue that AI can be used to mine data in electronic health records (EHR) in order identify the preferences of people who are otherwise incapacitated, regarding healthcare decisions. The authors distinguish between decision-making capacity and incapacity (Lamanna and Lauren, 2018:903). A person who displays decision-making capacity can understand the information related to the decision and appreciate its significance, to weigh the benefits and costs of different alternatives and to communicate the choice that has been made. Within the medical community incapacity is a challenging problem. More than 30% of psychiatric hospital inpatients and the elderly lack decision-making capacity and medical professionals have failed to identify incapacity in more than 40% of cases. The result is a failure to consider patients preferences when designing a treatment plan.

Using patient surrogates or family members to articulate preferences for incapacitated patients is a limited solution. They wrongly predict preferences 30% of the time and normally transfer their own preferences to the patient. Further this causes many surrogates to experience mental health problems and stress (Lamanna and Lauren, 2018:904).

Utilizing AI (Lamanna and Lauren, 2018:904) to predict preferences based on the entire patient population is equally accurate to using surrogates, if relevant information is available. The authors propose that by building a regression model including relevant factors such as age and marital status, the AI could become more accurate than surrogates in areas such as cancer treatment.

The authors (Lamanna and Lauren, 2018:905) describe two challenges to overcome to revolutionize predicting preferences related to health care decisions through machine learning. First is to provide AI with datasets containing population-wide electronic health records and second is to structure the information so that it is interpretable by AI (machine readable). While this is assumed to lead to improvements in predicting preferences the authors would like to go further and incorporate political and religious preferences as well as risk tolerance by examining a patient’s social media profiles to further improve accuracy.

The AI approach to mining preferences would not only lead to improved confidence in preference accuracy but could additionally reduce the burden of making life-or-death decisions currently carried by surrogates (Lamanna and Lauren, 2018:906).
The authors note that this novel approach does present some problems. The AI may just reproduce existing biases and reinforce prejudices instead of reflecting a patient’s true preferences. Another potential issue could be if a patient is incorrectly diagnosed as lacking decision-making capacity and the AI is used to override their actual preferences (Lamanna and Lauren, 2018:907).

An advantage could be that for patients with decision-making capacity and who are correctly diagnosed as such, the AI could act as a decision aid (Lamanna and Lauren, 2018:907). But the question remains how a patient should choose if the confidence levels of the AI and their doctor differ? Should they defer to their doctor, even though the AI could present a more rationally accurate reflection of their preferences? Overall the authors (Lamanna and Lauren, 2018:908) conclude that for an AI to provide benefits when used in this capacity, it should support people in making decisions rather than acting truly autonomously. Real (as opposed to theoretical) life or death decision situations are an area not commonly explored, perhaps unsurprisingly, by other researchers.

### 2.3.6 Fear of robots and AI

Liang and Lee (2017:379) studied the extent to which people have negative reactions to AI and autonomous robots. They propose a concept called "fear of autonomous robots and artificial intelligence" (FARAI) as a way of understanding people’s perceptions of human–robot interactions (HRI). The authors note that AI and autonomous robots are technically and conceptually different but that the public’s response to either is indistinguishable. Liang and Lee (2017:379-381) conducted a survey (large sample, n>1500) to elicit the attributes of people who experience FARAI. The authors posed three questions: how prevalent is fear towards autonomous robots in the US population, how do demographic variables impact this fear and what is the relationship between fear of robots and exposure to science fiction? The participants were asked questions relating to their level of fear regarding robots acting and deciding on their own, robot workers replacing people, “AI” as a general term/topic and trusting AI to do work.

Regarding the prevalence of FARAI, the researchers (Liang and Lee, 2017:382-383) found that roughly 25% of the respondents were fearful. Several weak predictors were found to correlate with increased FARAI including higher age, lower education and lower income and if the respondent was a woman. There was additionally a slight tendency towards increased FARAI for individuals with significant exposure to science fiction. FARAI was shown to positively correlate with other fears such as loneliness and unemployment. The authors note that even though most people have yet to experience AI or autonomous robots, a significant portion of the US population is already fearful of them. The demographic results point to that fear must be considered when AI or autonomous robots are implemented in these populations.
(and in general). Another finding is that increased media exposure may be a strategy for reducing FARAI by painting positive pictures of autonomous robots and AI. The positive correlation to loneliness suggests a fear of decreased human interaction and social displacement and the correlations to unemployment suggests a fear of losing work to robots. The authors conclude that

"No empirical research has so far documented an actual case of robophobia; however, as robots and artificial intelligence permeate wide segments of daily life, we can speculate its emergence in the future." (Liang and Lee, 2017:384)

2.3.7 Negative attitudes towards robots

Nomura et al. (2006:138) investigated negative attitudes towards communication robots as a factor preventing humans from interacting with them, since earlier research has considered computer anxiety as a barrier to learning in educational environments. The authors (Nomura et al., 2006:139) developed a measure called Negative Attitude toward Robots Scale (NARS) to evaluate psychological factors hindering people from interacting with robots in daily life. They conducted an experiment with a communication robot called "Robovie" in order to study negative attitudes towards robots.

Participants (n=53) in the experiment were asked to complete a survey of NARS related questions and requested to interact verbally and physically with "Robovie" (Nomura et al., 2006:141-142). The behaviors of the participants together with the robot were recorded using digital video cameras and closeness together with how much time elapsed before humans responded to requests for verbal or physical interaction by the robot, were measured.

The authors (Nomura et al., 2006:147) results suggests that a negative attitude towards robots has an impact on human behavior (avoidance), that there are gender differences in negative attitudes towards robots (women being less negative) and that peoples real experience with robots has an impact on their negative attitude (slower response time) and behavior (larger distance). They conclude that gender and previous experience with robots should be considered when designing the behavior and appearance of robots. As this study was conducted in Japan the researchers highlight that fact that the findings may not generalize to other cultures (Nomura et al., 2006:147). As the researchers did subject the participants to a situation in which anxiety towards robots would cause discomfort, the decision to enter the room did constitute somewhat of a risk for the participants.

2.3.8 AI and job elimination

Broughham and Haar (2018:239) studied a concept called Smart Technology, Artificial Intelligence, Robotics, and Algorithms (STARA) in relation to how these technologies together could potentially eliminate up to 30% of jobs by 2025. The au-
thors wanted to understand how employees perceived the impact of these technologies on their jobs and careers and how they were preparing for the change. The authors (Broughham and Haar, 2018:240) note that STARA could impact and displace up to 47% of all jobs, and not only low-skilled, low-paid ones but also occupations within the finance, medical, education, transportation, farming and service industries. The purpose of the study was to ascertain how to capture STARA awareness, test employee awareness and impact of STARA and the effect of STARA on a range of well-being and job outcomes. They predicted that STARA awareness would be negatively correlated with organizational commitment and career satisfaction but positively correlated with turnover intentions, depression and cynicism (Broughham and Haar, 2018:242-243). Additionally, Broughham and Haar (2018:244) thought that STARA awareness would be higher for younger employees and have a more detrimental effect on outcomes. The authors combined paper-based and online surveys (n=120) to gather data.

The researchers found that STARA awareness was low but that it was correlated as expected with organizational commitment, career satisfaction, turnover intentions, depression and cynicism as well as age (Broughham and Haar, 2018:245). STARA was not correlated to job security. The authors' interpretations of the results are that STARA awareness is overall low (mean result of 1.7 on a 5-point scale). They propose that STARA awareness may be low since it could be overshadowed by broader trends in the work environment such as lifetime employment being replaced by temporary contracts, driving overall lower employee commitment and loyalty (Broughham and Haar, 2018:252). Broughham and Haar (2018:253-254) advise employees to be mindful of the impact of STARA on their industries and plan career changes accordingly. At the same time the authors acknowledge that we in fact don't know if STARA is a net job creator or job destroyer because of this new industrial revolution.

2.3.9 Comparative advantage, AI vs. human

Jarrahi (2018:1) examined how humans and AI can utilize their different strengths in decision-making situations characterized by complexity, uncertainty and equivocality (different interpretations). The author (Jarrahi, 2018:2) describes several applications of AI including machine learning (learn from data and experience), machine vision (image processing) and natural language processing (understand human sentences) enabling AI to act as semi-autonomous decision-makers. The excitement and fear surrounding humans being replaced by AI is not a recent phenomenon. In 1965 Herbert Simon predicted machines could perform any work that a human could do by 1985. Similarly, in 1970 Marvin Minsky predicted that a machine with the general intelligence of an average human would exist within three to eight years (Jarrahi, 2018:4). Missing from that discourse, and the current one, is a discussion of
the different strengths and weaknesses of humans and AI. The starting point for Jarrahi’s (2018:4) exploration is that "computers plus humans do better than either one alone" and he looks at how humans and AI can complement one another in organizational decision-making.

An example of humans and AI performing better together comes from the study of cancer detection (Jarrahi, 2018:5). Humans alone performed better than AI alone, but by working together the error rate could be reduced by 85% compared to the human alone. In organizational decision-making the author distinguishes between analytical and intuitive decision-making and identifies three challenges: complexity, uncertainty and variation in interpretation (equivocality). Analytical decision-making requires information gathering and logical analysis including developing alternatives and consequences. In contrast to intuitive decision-making, the analytical sphere is an area in which tools such as DSS and predictive analytics can and do provide support in handling otherwise unmanageable amounts of data. Intuitive decision-making (a.k.a. "gut feeling") entails generating knowledge directly and deciding without relying on logical inference or rational thought, by drawing on past experiences, judgements and practices. This might be useful in ambiguous decision situations or where no precedent exists (Jarrahi, 2018:5-6).

Uncertainty is defined as a lack of information about the decision alternatives and their consequences and in an organizational context may refer to both the external and internal environment. In such a case AI may be able to generate new ideas using probabilistic approaches and identify otherwise unseen relationships between factors, allowing humans to act with the support of AI, even if information is limited. Simultaneously while AI can be employed in probability-based decision situations it is ill-equipped to deal with novel problems, i.e. the real world is not a chessboard. AI support systems may struggle with "common-sense situations" (judgement), compared to humans and have a hard time navigating uncertain situations, especially outside the domain of pre-defined knowledge (Jarrahi, 2018:6-7).

Complex situations are characterized by many variables or elements. They involve processing large amounts of information beyond the abilities of human decision-makers. Combined with big data and deep learning, machines can deal with complexity and find effective ways of providing data visualization and analytical support to humans. Causal relationships, invisible to humans, can be identified. To realize a synergic relationship between AI and human, the speed of data collection and information analyses can be combined with human’s superior intuitive insight and judgement (Jarrahi, 2018:8).

Equivocality is defined as the existence of several divergent and simultaneous interpretations of a decision. It is commonly encountered as a conflict of interest between different stakeholders. To fulfill the conflicting needs and preferences of mul-
tiple parties, rational and analytical thinking may be insufficient, and the only way forward might be through a subjective, political process. While AI possibly can support with sentiment analyses of conflicting needs, equivocality is ultimately the domain of human actors. Even if a machine could identify the optimal solution for all stakeholders, it would be hard-pressed to convince the parties involved to accept the decision. The ability to apply social and emotional intelligence to develop a vision and convince others to follow and achieve it (storytelling) requires putting interpersonal skills into action. If humans are to "follow the leader", he/she/it must be able to prove this ability (Jarrahi, 2018:8-9).

Organizational decisions are often characterized by uncertainty, complexity and equivocality requiring the combination of analytical and intuitive approaches to decision-making which necessitates combining the skills of humans and AI in a sort of human-machine symbiosis. While machines focus on mundane, repetitive tasks, humans are free to peruse and focus on creativity (Jarrahi, 2018:9-11). Overall Jarrahi (2018:6) suggests that

"decision-making will likely remain a comparative advantage for humans who can leverage superior intuition, imagination and creativity."

2.3.10 Risk perception and autonomous vehicles

Brell et al. (2018:1) studied risk perception towards driving connected and autonomous vehicles in comparison to conventional ones. The authors (Brell et al., 2018:2-3) incorporate both social perspectives of risk,

"the possibility of consequently affecting what a person values through human actions or events,"

as well as technological perspectives,

"the likelihood of physical, social, and/or financial harm/detriment/loss as a consequence of a technology aggregated over its entire life-cycle,"

into their work. Overall the authors state that the acceptance of technology is influenced by the perception of its risks and benefits. In relation to autonomous driving, one of the main fears is giving up control of the vehicle but other concerns include financial risks, social risks, ecological risks and the fear of hackers taking over.

To conduct their study Brell et al. (2018:4) looked at three types of driving technology: conventional driving connected driving and autonomous driving. The authors formulated two hypotheses, that perceived risks would differ depending on the driving technology and that prior experience with driver assistance systems would decrease the risk perception. Brell et al. (2018:5-6) conducted focus group research and executed a survey based on scenarios related to the three different types of driving technology. In general, the authors (Brell et al., 2018:9) found that
connected and autonomous driving was perceived to be more dangerous and costlier than conventional driving, but more beneficial regarding comfort and innovation. Additionally, those with high experience of driving aids ranked conventional vehicles as riskier than the other types on the categories of traffic environment, vehicle and passenger but the least risky when it came to the category of data. Those with less experience of driving aids overall ranked conventional vehicles to be less risky in all categories.

Overall the authors (Brell et al., 2018:12) found that more experience with driving aids increased the perceived benefits of connected and autonomous vehicles. Another finding was that there was little distinction between connected and autonomous vehicles in the perceptions of the participants but rather different views of new (connected, autonomous) and old (conventional) technologies. Additionally, even though conventional driving is factually more dangerous in terms of hazards and accidents, this "rational" perception wasn't shared by the participants in the study.

In conclusion Brell et al. (2018:14) describe a narrative where

"self-driving vehicles are perceived as spooky, mechanic, not controllable, and somehow immoral."

But at the same time increased experience with a technology does seem to have a potential positive impact from a risk perspective, resulting in a decrease in the overall perception of danger.

### 2.3.11 Ethical AI

Etzioni and Etzioni (2017:403) reviewed why it is believed that AI must have the capability of making ethical decisions and the challenges of this approach. Their focus is primarily on autonomous AI decision-making and related to applications such as driverless cars, self-targeting weapon systems or robotic surgeons. The authors differentiate between legal and personal choices. That a car, autonomous or conventional should make a full stop at a stop sign is an ethical decision made collectively through the normal process of law-making. Whether a human driver or autonomous car should stop to pick up a stranded motorist is an ethical decision at the individual level.

Since AI such as an autonomous vehicle, can cause harm, it is argued that there is a requirement for AI to be able to differentiate between "right and wrong" decisions (Etzioni and Etzioni, 2017:405). Therefore, some believe that AI should be "moral reasoners" capable of making their own ethically sound decisions. An example of an ethical dilemma that an AI should be expected to navigate is commonly an adaptation of the Trolley Problem, where an AI would be forced to choose between only morally repugnant actions such as injuring or killing different groups of people. The
problem with this scenario that it describes a rare, outlier case, not a type of situation an AI would normally encounter in the real world. Additionally, we accept that accidents occur and that "human error" is the cause (Etzioni and Etzioni, 2017:416). Is it reasonable to hold AI to a higher standard in extreme cases, when it would probably perform much better (safer) than humans on average?

The authors (Etzioni and Etzioni, 2017:410-411) differentiate between two types of AI, roughly corresponding to two definitions presented earlier in this paper (see chapter 2.1). Thinking human, i.e. replicating and replacing human thought and acting rational, i.e. agents providing "smarter" (than human) assistance to human decision-makers. They argue that it is only in cases where an AI truly acts independently and replaces a human decision-maker, with humans removed from the loop, that an AI would be required to have "moral autonomy."

Two approaches have been suggested for implementing ethics into AI. A top-down approach (Etzioni and Etzioni, 2017:405-406) would program an AI with ethical principles, for example the Ten Commandments, Asimov’s Three Laws of Robotics or some other general philosophy such as utilitarianism. The AI would be expected to make ethical choices, not follow a pre-programmed script for every situation. The main criticisms of this approach include the inability of estimating long-term consequences. Is utility even quantifiable and would it "lead to actions and outcomes that many will find morally unacceptable" (Etzioni and Etzioni, 2017:406)?

The bottom-up approach (Etzioni and Etzioni, 2017:406-407) in contrast expects machines to learn ethics from human role-models by observing their behavior, without being taught ethical rules or programmed with a moral philosophy. One way of achieving this could be by applying machine learning techniques. The challenge is that an AI "would have to follow a person for several lifetimes to learn ethics in this way" (Etzioni and Etzioni, 2017:407). To this end it might be possible to aggregate the decisions and behaviors of millions of people when teaching an AI, but the risk is that the AI learns what is common (i.e. cars speeding), not what is ethical (Etzioni and Etzioni, 2017:407).

Another assumption the authors challenge is the belief that AI is truly autonomous and therefore a "moral agent" (Etzioni and Etzioni, 2017:408-409). While it is taken for granted that humans have "free will", not everyone agrees that even humans are truly autonomous, as they are constantly subjected to forces outside of their control. The authors describe autonomy as existing on a continuum, from no autonomy to full autonomy. Since it is impossible to write code to anticipate every situation, a technique such as deep learning can be used to allow the machine to figure things out by itself, which may seem like autonomy, since how the AI makes decisions may be a mystery even to its creators. But the AI is still constrained by the human that
programmed it. The AI isn't "morally autonomous" as it can't challenge the situation by objecting to or changing its mission. As stated by Brad Templeton,

"a robot would be truly autonomous the day it is instructed to go to work and it instead goes to the beach." (Etzioni and Etzioni, 2017:409)

The current state of AI means that it could make some decisions autonomously, but to make ethical decisions it still requires support from a human.

According to Etzioni and Etzioni (2017:410) human autonomy can be violated, while AI autonomy cannot. An AI doesn't have emotions and it doesn't feel pain. There is no ethical objection to changing the programming of an AI to prevent it from harming others. This is in stark contrast to the moral implications of reprogramming a human. This may make it easier to help AIs "acquire moral values" as it "only" requires reprogramming with the correct constraints, compared to the re-education, shaming or possible imprisonment to right the misbehavior of a human. The proposal for the foreseeable future is a division of labor, humans partnering with AIs each responsible for their own comparative advantage, with humans providing moral guidance (Etzioni and Etzioni, 2017:411). One way of implementing this is to allow humans to configure AIs with their value preferences by providing option settings. This may not in fact be a way forward since many humans prefer to accept the default settings (Etzioni and Etzioni, 2017:413). An alternative could be an "ethics bot", an AI which would analyze information about an individual and deduce that person's moral preferences. This would of course entail the same pitfalls as using an AI for any other purpose. What happens if for example the individual is irresponsible or even a lawbreaker? The answer is to only allow the AI freedom regarding individual ethics but to continue to be constrained by collective ethics (laws, regulations, etc.).

2.3.12 Prospect theory and human “irrationality”

To create an AI that thinks like a human we must understand how humans think. Russell and Norvig (2010:636) describe the construction of an AI agent capable of selecting actions to maximize performance by combining probability theory, utility theory and decision theory. Such an AI agent would be “rational” and make rational decisions. Would such an AI agent accurately reflect the way that humans think? The research of Kahneman and Tversky (1979) suggests that this may not be the case.

Kahneman and Tversky (1979:263) conducted several studies and present a critique of expected utility theory related to decision making under risk and propose an alternative model called prospect theory. Instead of assigning value to the final state (sum of assets) together with probability as utility theory does, they suggest assigning value to the resulting gains or losses of a decision in conjunction with decision weights. Regarding human decision making the authors state that
“our perceptual apparatus is attuned to the evaluation of changes or differences rather than to the evaluation of absolute magnitudes.” (Kahneman and Tversky, 1979:277)

The authors identify several phenomena in which the tenets of expected utility theory are violated (Kahneman and Tversky, 1979:264). They identify the "certainty effect" showing that people tend to overweight outcomes (gains) that are considered certain in relation to outcomes which are only probable, which seems to suggest risk aversion to sure gains (Kahneman and Tversky, 1979:265). The “reflection effect” implies the inverse for situations related to sure losses where people instead tend towards risk seeking (Kahneman and Tversky, 1979:268). The isolation effect is a phenomenon identified by the authors where people tend to disregard components that are shared between alternatives and instead only focus on those parts that differ. The result is that people make inconsistent decisions when the same choice is presented in different ways (Kahneman and Tversky, 1979:271).

Prospect theory was developed because of the observed effects in relation to simple decisions with given probabilities and monetary consequences, but the authors thought the results could be generalized to other areas (Kahneman and Tversky, 1979:274).

2.4 Summary and relation to hypotheses

The literature review has provided an overview of the areas of artificial intelligence, decision-making and risk. Previous research has focused on perspectives such as ethics, trust, fear and negative attitudes in relation to decision-making and artificial intelligence investigated along dimensions such as gender and age. The literature review was used to guide and inspire the formulation of the hypotheses to be investigated. Primarily the different perspectives related to decision-making and artificial intelligence in section 2.3 were used.

The study by Brell et al (2018:1) on perceived risk and autonomous driving as well as the Aroyo et al. (2018:3702) study on social engineering including risk aversion inspired H1. H2 was guided by the work of Aroyo et al. (2018:3702), Liang and Lee (2017:379) and Nomura et al. (2006:138) regarding negativity towards AI and robots. H3, H4 and H5 were inspired by the work of Lee (2018:4-5) regarding AI fairness and trust as well as the work by Dietvorst et al. (2014:2) on algorithm aversion. Liang and Lee (2017) focused on a US population and Nomura et al. (2006) on Japan. The hypothesis H6A was chosen to align with Liang and Lee’s (2017) findings as this author thinks the US is a more heterogeneous culture, better reflecting the sample chosen for this study. H6B was inspired by Liang and Lee (2017:379) investigating the relationship between fear of AI and age.
3 Design of study

The purpose of this chapter is to describe the design of the study and the implementation of the chosen method and includes perspectives on research ethics.

The chosen area of study was attitudes to artificial intelligence decision-making support under risk. The chosen level of analysis was the individual decision maker relying on human or AI decision support in situations with varying risk. Several hypotheses were formulated to better articulate the question to be studied. To conduct this study a survey was executed with a mixed quantitative and qualitative approach. One open, purely qualitative question was included to enable participants to share their viewpoints and capture perspectives that might otherwise have been disregarded.

The chosen area of study focused on individual decision makers who themselves subjectively and interactively defined the content and scope of the decision situations under consideration including their personal definitions of AI and risk. The investigation was comparative, contrasting human vs. AI decision support and high-risk vs. low-risk decision situations. Consequently, the strategy for investigation was chosen because the author thinks that it was best suited to answer the question of the study and to test the hypotheses.

3.1 Method

The chosen method was to conduct a literature review combined with a quantitatively focused survey with one qualitative question intended to potentially test the pre-formulated hypotheses and answering the question of “How are people’s attitudes to artificial intelligence vs. human decision-making support affected by the level of perceived risk in the decision situation?”

3.1.1 Chosen risky scenarios

Participants were presented with two theoretical scenarios with varying degrees of risk (thought experiments) describing situations that this author supposes they may have already been subjected to or in the future were likely to encounter. The two risky decisions situations are summarized in Table 2 below.

<table>
<thead>
<tr>
<th>Risk level</th>
<th>Decision scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-risk</td>
<td>Purchase a $500,000 home</td>
</tr>
<tr>
<td>Low-risk</td>
<td>Select restaurant for $50 dinner</td>
</tr>
</tbody>
</table>

The high-risk situation presented to participants was to purchase a new home valued at $500,000 USD. The low risk situation was described as the selection of a restau-
rant in order to buy a dinner valued at $50 USD. Both scenarios described that the participants were requesting help in making the decision and that as a decision-maker they were trying to maximize value, according to their own definition and understanding of value.

Personal decision situations with a high likelihood of personal stakes were chosen instead of for example business decisions where a loss could be shared (company loses money and decision maker may lose status, job, etc.). Additionally, the perceived risk in business scenarios is often informed by the level of experience and expertise in an area (i.e. marketing, production, purchasing, etc.) which could have created challenges when calibrating perceived risk between participants. To test the hypotheses the participants were asked different questions to evaluate their willingness to rely on and accept the advice of humans vs. machines to support them in making decisions in situations of varying risk.

3.1.2 Data collection and survey structure

The survey was created in an online survey tool (https://sv.surveymonkey.com/) to collect data. The survey consisted of 5 parts spread over 7 pages including 24 questions. An example of the online survey tool can be seen in Figure 2 below. The survey itself can be found in Appendix A.

Part 0 consisted of a welcome page describing the purpose of the investigation and the fact that participants would be asked to evaluate AI and human decision-making support in different risky scenarios. The number of questions and time estimated to complete the survey was stated. An e-mail contact was presented so that participants could raise questions or voice concerns. The survey was presented as anonymous and that online data would only be retained for one month. Offline data was stated to be retained until the study was completed. The online survey tool was configured to not capture IP addresses to provide anonymity.
Part 1 and 2 consisted of six questions on one page to gage the risk tolerance of the participants as well as their overall attitudes to AI and human decision-making support.

Guillemette et al. (2015:17) evaluated how well different questions for assessing risk correlated with actual monetary loss aversion based on utility theory, prospect theory and self-assessment. Consequently, the risk assessment instrument developed by Grable and Lytton (2003:271) was identified as a suitable inspiration for formulating four questions designed to divide the survey participants into risk seeking and risk averse groups for analysis (See questions 1-4 in Appendix A). To understand positive or negative attitudes to AI decision-making support a five-level question was developed based on the work of Nomura et al. (2006:140). A similar question was then formulated to evaluate human decision-making support. This data was used to further divide the participants into groups for analysis.

Part 3 covered the home purchasing scenario (high risk) spread over two pages and six questions. The scenario was described to the participant including the potential good or bad outcomes. Two questions were asked to allow the participant to evaluate the perceived risk and impact of the decision situation. Perceived risk was captured on a numeric scale of 0-100, 100 representing very high perceived risk and 0 indicating no perceived risk. The overall impact of the decision situation to the participants’ life was graded from very small impact to very large impact on a five-level Likert scale.

On the following page of the survey the participants were instructed to provide their attitudes to AI and human decision-making support in the high-risk scenario. The AI decision-making support was presented as a self-service real estate website while the human decision-making support was described as a real estate agent. The recommendations from either could span a range from a bad home to a good home. The participants were asked to evaluate the reliability of the recommendation provided by the AI and human and how likely they would be to accept the recommendation of either type of decision support on a five-level Likert scale from very unlikely to very likely. Acceptance was chosen as a stronger alternative to trust (compare Dietvorst et al., 2014), as it requires a subject to both trust and act on that trust.

Part 4 consisted of the restaurant selection scenario (low risk) containing six questions over two pages. The scenario description and guiding texts were like the high-risk scenario and the questions were identical. The decision-making support for AI was described as a self-service restaurant recommendation website and the human support was presented as a concierge (reception worker). The outcomes ranged from a bad dinner to a good dinner.
Part 5 was presented on the final page and contained six coding and additional questions. The participants were asked to share their age from six age ranges, their gender (male/female/other/unsure/no answer, via RFSL, 2016), the country they were currently residing in and their current work status. A yes or no question was presented to ascertain if the participants had ever used AI in support of a real decision, according to the best of their knowledge. The final question was an open question allowing the participants to describe any general thoughts or feelings they may have regarding utilizing AI to provide decision-making support.

3.1.3 Sampling

Participants were recruited via the author’s global, professional network, mainly using LinkedIn.com. The participants were selected via non-random judgement sampling and the author chose participants subjectively, intending to ensure a broad selection of participants from a wide variety of backgrounds. Contact with the participants was conducted via LinkedIn.com. Information about the purpose of the study was provided to the participants including a link to the online survey tool.

3.1.4 Data analysis

Data collected according to Likert scales was forward-scored on a numeric scale of 1-5, for example: "Very Unlikely" = 1; "Unlikely" = 2; "Neither likely nor unlikely" = 3; "Likely" = 4 and "Very Likely" = 5 to enable statistical analysis. When presenting results and conducting analysis data such as “Very Unlikely” and “Unlikely” were grouped together as “Unlikely” and data such as “Likely” and “Very Likely” were grouped together as “Likely”. Percentages were then calculated for each grouping. Data on the scale of “Very Negative” to “Very Positive” and “Very Small Impact” to “Very Large Impact” were grouped and calculated in the same manner.

The data collected according to Likert scales was assessed using a Mann-Whitney test (Laerd Statistics, 2018a) for groups with different participants (i.e. male or female) and a Wilcoxon signed-rank test (Laerd Statistics, 2018b) when the same participants were present in both groups (i.e. same participants in multiple test conditions). Two nonparametric statistical tests were chosen due to the data being discrete, ordinal, limited in range and consisting of relatively small sample sizes. Tests are two-tailed unless otherwise specifically mentioned.

The tools used for analysis were Excel and Minitab 14.

3.2 Research ethics

The research used two simulated decision situations (thought experiments) under risk when conducting the survey. The study as such was therefore not sensitive in nature. The survey participants were explicitly notified that they would be anony-
mous, and the results were anonymized. Neither personal data nor IP addresses were captured or retained. An anonymity agreement wasn’t created. Participants came from many backgrounds and the results can’t be tied to any one organization. Based on these facts it was decided by this author and his advisor that this study didn’t need to be approved by the university ethics board.
4 Survey results

The purpose of this chapter is to present the results of the online survey. An analysis of the results is presented in chapter 5.

4.1 Survey sampling and response

The online survey was sent to 412 people and was open from October 24th, 2018 to November 8th, 2018. 123 people responded to the survey for an overall response rate of 30%. In total 16 people started the survey but didn’t complete it. A total of 107 people completed all non-optional questions of the survey for a response rate of 26%. The optional qualitative question was completed by 47 participants.

4.2 Participant coding

Most participants (85%) were between the ages of 35 to 54. Only 2 participants were under age 35 and 14 were above age 54. Most participants were male (70%) and a minority were female (28%). Most participants were residing in Sweden (75%) and a few were in the US (10%). Participants were in 11 countries total but no other country, besides Sweden and the US, contained over 4% of participants. The work status for 93% of the participants was working (part-time, full-time or self-employed).

The typical participant in this study is male, age 35-54, residing in Sweden and currently employed in some manner. It is quite likely that this participant has previous experience with AI decision-making, since out of all participants, 44% answered they had this experience.

A total of 107 participants answered the age question; see Figure 3 for the age distribution of the participants:

![Age distribution](image)

*Figure 3 — Age distribution*
A total of 107 participants answered the gender question; see Figure 4 for the gender distribution of the participants:

![Figure 4 – Gender distribution](image)

A total of 105 participants answered the country of residence question and as shown in Figure 5 the participants resided in the following countries when taking the survey:

![Figure 5 – Country of residence](image)

A total of 107 participants answered the work status question; see Figure 6 for the distribution of the participants:
A total of 47 out of 107 participants answered that they had used an AI to assist them in making a real decision to the best of their knowledge, indicating a minority of 44%.

### 4.3 Attitudes to risk

A total of 123 participants answered the four risk related questions: 1) “When you think of the word "risk" which of the following words comes to mind first?”, 2) “You are on a TV game show and can choose one of the following. Which would you take?” 3) “In general, how would your best friend describe you as a risk taker?” and 4) “You have just finished saving for a “once-in-a-lifetime” vacation. Three weeks before you plan to leave, you lose your job. What would you do?” Each question had four alternatives the participants could choose from, see <Page 2> in Appendix A.

According to Grable and Lytton (2003:272-273) the answers to the risk perception questions can be scored on a scale of 1-4. The scores 1-2 are interpreted as risk averse and 3-4 interpreted as risk seeking per question. This scoring system was used to compile a total risk score for each participant. Each risk question was scored 1-4 and the total risk score was calculated from the scores of each of the four questions. A total risk score of 4-8 was interpreted as risk averse and 9-16 as risk seeking. The scoring model was constructed according to the methodology described by Grable and Lytton (2003) regarding how attitudes to risk should be classified. Consequently, there is an unequal division between the value distribution of risk aversion (5 values, 4-8) and seeking risk (8 values, 9-16) in the total risk score. The results of the scoring are summarized below in Figure 7:
Overall 16% of participants were identified as risk averse based on their total risk score. The average risk score was 9.76 and the median was 10.

### 4.4 Attitudes to decision support

A total of 123 participants answered the question “What is your overall attitude to utilizing AI (artificial intelligence) to help you in your decision-making?” and the responses are shown in Figure 8:

As a result, 64% of participants were overall positive towards receiving decision-making support from AI and 3% were overall negative. Data was grouped according to the description given in 3.1.4.

A total of 123 participants answered the question “What is your overall attitude to utilizing other people to help you in your decision-making?” and the responses are shown in Figure 9:
As a result, 84% of participants were overall positive towards receiving decision-making support from humans and 3% were overall negative.

### 4.5 Home purchasing scenario results

A total of 113 participants responded to all six questions in the high-risk scenario.

The distribution of answers to the question “How high is the perceived risk for you when purchasing a new home valued at $500,000 USD?” was as follows shown in Figure 10:

The average perceived risk was 43.4. 65% of participants perceived the risk to be low (0-50) and 35% of participants perceived the risk to be high (51-100). Data was grouped according to the description given in 3.1.4.

The distribution of answers to the question “How do you think this home purchase will impact your life overall?” was as follows in Figure 11:
As a result, 72% of participants thought there would be a large impact on their lives and 9% thought there would be a small impact on their lives in the high-risk scenario. Data was grouped according to the description given in 3.1.4.

The distribution of answers to the question “How likely do you think it is that the AI (artificial intelligence) will suggest a recommendation that you can rely on?” was as follows in Figure 12:

As a result, 65% of participants were likely to rely on the AI recommendation and 8% of participants were unlikely to rely on the AI recommendation in the high-risk scenario. Data was grouped according to the description given in 3.1.4.

The distribution of answers to the question “How likely are you to accept the home recommended by the AI (artificial intelligence)?” was as follows in Figure 13:
As a result, 41% of participants were likely to accept the AI recommendation and 8% of participants were unlikely to accept the AI recommendation in the high-risk scenario.

The distribution of answers to the question “How likely do you think it is that the real estate agent will suggest a recommendation that you can rely on?” was as follows in Figure 14:

As a result, 53% of participants were likely to rely on the human recommendation and 12% of participants were unlikely to rely on the human recommendation in the high-risk scenario.

The distribution of answers to the question “How likely are you to accept the home recommended by the real estate agent?” was as follows in Figure 15:
As a result, 39% of participants were likely to accept the human recommendation and 7% of participants were unlikely to accept the human recommendation in the high-risk scenario.

### 4.6 Restaurant selection scenario results

A total of 110 participants responded to all six questions in the low risk scenario.

The distribution of answers to the question “How high is the perceived risk for you when choosing a restaurant to have a dinner valued at $50 USD?” was as follows in Figure 16:

The average perceived risk was 23.6. 86% of participants perceived the risk to be low (0-50) and 14% of participants perceived the risk to be high (51-100).

The distribution of answers to the question “How do you think this dinner will impact your life overall?” was as follows in Figure 17:
None of the participants thought there would be a large impact to their lives and 94% thought there would be a small impact to their lives in the low-risk scenario.

The distribution of answers to the question “How likely do you think it is that the AI (artificial intelligence) will suggest a recommendation that you can rely on?” was as follows in Figure 18:

As a result, 81% of participants were likely to rely on the AI recommendation and 3% participants were unlikely to rely on the AI recommendation in the low-risk scenario.

The distribution of answers to the question “How likely are you to accept the dinner recommended by the AI (artificial intelligence)?” was as follows in Figure 19:
As a result, 84% of participants were likely to accept the AI recommendation and 2% of participants were unlikely to accept the AI recommendation in the low-risk scenario.

The distribution of answers to the question “How likely do you think it is that the concierge will suggest a recommendation that you can rely on?” was as follows in Figure 20:

As a result, 68% of participants were likely to rely on the human recommendation and 5% of participants were unlikely to rely on the human recommendation in the low-risk scenario.

The distribution of answers to the question “How likely are you to accept the dinner recommended by the concierge?” was as follows in Figure 21:
As a result, 71% of participants were likely to accept the human recommendation and 5% of participants were unlikely to accept the human recommendation in the low-risk scenario.

### 4.7 Qualitative data

The open, qualitative question asked participants to “please share any thoughts or feelings you may have about using AI (artificial intelligence) to assist people in making decisions”. A total of 42% of comments were focused on attitudes towards AI and overall twice as many comments (12 vs. 6) were positive to AI. A total of 26% of comments were focused on perspectives of trust in relation to AI. A total of 21% of comments were regarding the availability and quality of data and information. A total of 19% of comments highlighted aspects of rational AI and emotional humans. Out of the 47 comments, 4 were non-comments, for example “N/A”.

As mentioned, a total of 12 of the comments focused on the positive aspects of AI, for example:

“All over the interview I have assumed that AI is what people normally call ”AI” (algorithmic machine learning). Real Artificial Intelligence (non-algorithmic machine learning) is something completely different that doesn’t exist yet, or can be said to be at a very primitive stage. Real AI (whenever we get it) will definitely assist in making decisions. Current ”AI” can help in simple cases (like the examples mentioned above, choosing restaurants, homes, and all that simple stuff), but not in critical decisions - excepted for those depending on a limited and measurable number of factors.”

In contrast some of the comments were clearly negative towards AI such as the following:

“"I believe that AI will take a big part in the daily life in the near future, not only private life but also in the professional life, to assist people making decisions. However I don’t believe in using AI that makes the decisions.”
A total of 11 of the comments focused on perspectives of trust in relation to AI, for example:

"If there is no risk of significant loss involved, then relying on AI is not something to worry about. However, if there is indeed a risk of significant loss involved, then own research would be complementing any AI suggestion - at least the first time. One would want to gain some trust before casually trust AI in such circumstances. It is also worth considering the impact for an individual, versus decision making for a larger group where the majority will benefit, but some will perceive a loss."

Thoughts regarding data and information, such as its availability and quality were on the minds of 9 participants, for example:

"From a personal perspective (and reflected in my answers) the problem is not so much the probability that the AI will give a good recommendation based on my preferences but rather my (in)ability to express my preferences clear enough. Then no amount of data or statistical analysis will help (unless of course also adding info on what the users actually want when expressing something else). As of now, I still believe that for example a skilled human realtor has experience that can help him/her to interpret my wishes beyond my expressed wishes (and maybe even my conscious wishes). But maybe not for much longer…"

Additionally, the question of rational AI and emotional humans was found in the comments of 8 participants, such as:

“At a higher risk at stake I tend to distrust AI which is a clear emotional reaction and not a rational one.”
5 Analysis

The purpose of this chapter is to analyze the data presented in the empirical presentation.

5.1 Participant overall risk evaluation of scenarios

Two decision scenarios were designed by this author (see chapter 3.1), purchasing a home and choosing a restaurant for a dinner. The home purchasing scenario was designed to be perceived as high-risk by the participants and the restaurant selection scenario as low-risk. For the home purchasing scenario 35% of participants answered that the scenario has high risk, with an average score by all participants of 43 on a scale of 0-100. For the restaurant selection scenario 86% of participants answered that the scenario has low risk, with an average score by all participants of 23 on a scale of 0-100. This confirms that most participants thought that the restaurant selection scenario was in fact low risk. For the home purchasing scenario 65% of participants answered it was low-risk, meaning they did not think it to be a very risky situation. Overall the average score for the home purchasing scenario (43) was almost double that of the restaurant selection scenario (23). This confirms that overall the participants thought the home purchasing scenario was significantly riskier than the restaurant selection scenario.

Regarding impact 72% of participants thought that the home purchasing scenario would have a high impact on their lives. Conversely 94% of participants answered that the restaurant selection scenario would have a low impact on their lives. Taken together as shown in Table 3, the perceived risk and impact confirm that the participants thought that the overall risk for the home purchasing scenario was in fact high and the restaurant selection scenario was indeed low.

Table 3. Per scenario participant grouping with perceived risk and impact

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Participant grouping</th>
<th>Perceived risk</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home purchasing</td>
<td>High</td>
<td>35%</td>
<td>72%</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>65%</td>
<td>9%</td>
</tr>
<tr>
<td>Restaurant selection</td>
<td>High</td>
<td>14%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>86%</td>
<td>94%</td>
</tr>
</tbody>
</table>

5.2 Risk aversion, reliability and acceptance

The first hypothesis (H1) was that risk averse participants would be less likely to rely on or accept an AI recommendation, as opposed to risk seeking participants.

A Mann-Whitney U test was used to determine if there was a statistically significant median difference between risk averse and risk seeking participants regarding relia-
bility and acceptance in the two scenarios. When comparing risk averse (n=19) and risk seeking (n=94) participants in the home purchasing scenario regarding reliability (p (adjusted for ties) = 0.9159) and acceptance (p (adjusted for ties) = 0.3899) no significant difference was found. When analyzing risk averse (n=19) compared to risk seeking (n=91) participants in the restaurant selection scenario regarding reliability (p (adjusted for ties) = 0.7995) and acceptance (p (adjusted for ties) = 0.4669) the result was no significant difference between the groups found (Table 4).

The results in this study did not give support for H1.

Table 4. Participant risk score grouping and willingness to rely-on and accept AI recommendations per scenario

<table>
<thead>
<tr>
<th>Tested Variables</th>
<th>p (adjusted for ties)</th>
<th>Significant</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Risk averse vs. risk seeking</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliability</td>
<td>AI home purchasing</td>
<td>p = 0.9159</td>
</tr>
<tr>
<td></td>
<td>AI restaurant selection</td>
<td>p = 0.7995</td>
</tr>
<tr>
<td>Acceptance</td>
<td>AI home purchasing</td>
<td>p = 0.3899</td>
</tr>
<tr>
<td></td>
<td>AI restaurant selection</td>
<td>p = 0.4669</td>
</tr>
</tbody>
</table>

### 5.3 Negative attitudes to AI, reliability and acceptance

The second hypothesis (H2) was that participants with a negative attitude to AI would be less likely to rely on or accept an AI recommendation, as opposed to neutral or positive participants.

The intention was to use a Mann-Whitney U test to determine if there was a statistically significant median difference between participants with a negative attitude vs. a positive attitude to AI regarding reliability and acceptance in the two scenarios. It wasn’t possible to analyze the negative (n=3) and other (n=110) participants in the home purchasing scenario regarding reliability and acceptance. The sample size of negative participants was insufficient. The same problem was observed for negative (n=3) compared to other (n=107) participants in the restaurant selection scenario regarding reliability. It wasn’t possible to complete a statistical test since all negative responses were the same (all ties eliminated in the test).

Hypothesis H2 could not be tested.

### 5.4 Perceived risk and AI reliability

The third hypothesis (H3) consisted of two sub-hypotheses; participants will be more likely to rely on human rather than AI support in a decision situation characterized by high perceived risk (H3A) and participants will be equally likely to rely
on human and AI support in a decision situation characterized by low perceived risk (H3B).

The data was analyzed based on the participants own perceived risk for each scenario. For example, for the home purchasing scenario, defined as high-risk for the purposes of this study, some participants judged it to be high in perceived risk (n=40) and some thought it was low (n=73).

A Wilcoxon signed-rank test was used to determine if there was a statistically significant median difference between AI vs. human reliability based on perceived high and low risk in the two scenarios. When comparing reliability of AI vs. human based on perceived high risk in the home purchasing scenario (n=40, n for test = 26, p = 0.849) and the restaurant selection scenario (n=15, n for test = 5, p = 0.225) no significant difference was found (Table 5).

The results in this study did not give support for H3A.

When analyzing reliability of AI vs. human based on perceived low risk in the home purchasing scenario (n=73, n for test = 33, p = 0.081) and the restaurant selection scenario (n=95, n for test = 49, p = 0.033) the result was found to be a significant difference for the restaurant selection scenario but not the home purchasing scenario (Table 5).

The results in this study didn’t find acceptance for H3B and partially refuted it, finding that more participants found the AI to be reliable compared to the human in the restaurant selection scenario.

<table>
<thead>
<tr>
<th>Tested Variables</th>
<th>p</th>
<th>Significant</th>
<th>n for test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Risk</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI vs. human home purchasing</td>
<td>p = 0.081</td>
<td>No</td>
<td>33</td>
</tr>
<tr>
<td>AI vs. human restaurant selection</td>
<td>p = 0.033</td>
<td>Yes</td>
<td>49</td>
</tr>
<tr>
<td>High Risk</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI vs. human home purchasing</td>
<td>p = 0.849</td>
<td>No</td>
<td>26</td>
</tr>
<tr>
<td>AI vs. human restaurant selection</td>
<td>p = 0.225</td>
<td>No</td>
<td>5</td>
</tr>
</tbody>
</table>

5.5 Impact and AI reliability

The fourth hypothesis (H4) consisted of two sub-hypotheses; participants will be more likely to rely on human rather than AI support in a decision situation characterized by high impact (H4A) and participants will be equally likely to rely on human and AI support in a decision situation characterized by low impact (H4B).

The data was analyzed based on the participants own evaluation of impact for each scenario. For example, for the home purchasing scenario, defined as high-risk for
the purposes of this study, most participants judged it to be large in impact (n=81) and several thought it was small (n=10).

A Wilcoxon signed-rank test was used to determine if there was a statistically significant median difference between AI vs. human reliability based on evaluated large and small impact in the two scenarios. When comparing reliability of AI vs. human based on evaluated high impact in the home purchasing scenario (n=81, n for test = 46, p = 0.448) no significant difference was found (Table 6).

The results in this study did not give support for H4A.

When analyzing reliability of AI vs. human based on evaluated small impact in the restaurant selection scenario (n=103, n for test = 49, p = 0.018) the result was found to be a significant difference and in the home purchasing scenario (n=10, n for test = 5, p = 0.345) the result was no significant difference (Table 6).

The results in this study didn’t find acceptance for H4B and partially refuted it, finding that more participants found the AI to be reliable compared to the human in the restaurant selection scenario.

Table 6. Participant evaluated impact grouping and AI vs. human reliability per scenario

<table>
<thead>
<tr>
<th>Tested Variables</th>
<th>p</th>
<th>Significant</th>
<th>n for test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small impact</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI vs. human home purchasing</td>
<td>p = 0.345</td>
<td>No</td>
<td>5</td>
</tr>
<tr>
<td>AI vs. human restaurant selection</td>
<td>p = 0.018</td>
<td>Yes</td>
<td>49</td>
</tr>
<tr>
<td>Large impact</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI vs. human home purchasing</td>
<td>p = 0.448</td>
<td>No</td>
<td>46</td>
</tr>
</tbody>
</table>

5.6 Acceptance of AI recommendations

The fifth hypothesis (H5) stated that participants will overall be more likely to accept a human than an AI recommendation.

A Wilcoxon signed-rank test was used to determine if there was a statistically significant median difference between AI vs. human acceptance in the two scenarios. When comparing acceptance of AI vs. human based in the home purchasing scenario (n=113, n for test = 55, p = 0.990) no significant difference was found. The data indicates that the participants could be equally likely to accept the AI vs. human recommendation.

When analyzing acceptance of AI vs. human in the restaurant selection scenario (n=110, n for test = 56, p = 0.033) the result was found to be a significant difference. The data indicates that the participants were more likely to accept the AI vs. human recommendation (Table 7).
The results in this study didn’t find acceptance for H5 and partially refuted it, indicating that participants were potentially equally likely to accept AI and human in the home purchasing scenario and finding that more participants were willing to accept the AI compared to the human in the restaurant selection scenario.

Table 7. AI vs. human overall participant acceptance per scenario

<table>
<thead>
<tr>
<th>Tested Variables</th>
<th>p</th>
<th>Significant</th>
<th>n for test</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI vs. human home purchasing</td>
<td>p = 0.990</td>
<td>No</td>
<td>55</td>
</tr>
<tr>
<td>AI vs. human restaurant selection</td>
<td>p = 0.033</td>
<td>Yes</td>
<td>56</td>
</tr>
</tbody>
</table>

5.7 Gender and age vs. reliability and acceptance

The final hypothesis (H6) consisted of two sub-hypotheses; Women will be less likely to rely on or accept an AI recommendation (H6A) compared to men and older age groups (55+) will be less likely to rely on or accept an AI recommendation (H6B) compared to other age groups.

A Mann-Whitney U test was used to determine if there was a statistically significant median difference between female and male participants regarding reliability and acceptance in the two scenarios as well as older age groups (55+) compared to younger participants.

When comparing female (n=30) and male (n=83) participants in the home purchasing scenario regarding reliability (p (adjusted for ties) = 0.7742) and acceptance (p (adjusted for ties) = 0.9282), no significant difference was found. When analyzing female (n=30) compared to male (n=80) participants in the restaurant selection scenario regarding reliability (p (adjusted for ties) = 0.1332) and acceptance (p (adjusted for ties) = 0.6293), no significant difference was found (Table 8).

The results in this study did not give support for H6A.

Table 8. Female vs. other, per scenario, AI reliability and acceptance

<table>
<thead>
<tr>
<th>Tested Variables</th>
<th>p (adjusted for ties)</th>
<th>Significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female vs. other</td>
<td>Home purchasing</td>
<td></td>
</tr>
<tr>
<td>AI reliability</td>
<td>p = 0.7742</td>
<td>No</td>
</tr>
<tr>
<td>AI acceptance</td>
<td>p = 0.9282</td>
<td>No</td>
</tr>
<tr>
<td>Restaurant selection</td>
<td>AI reliability</td>
<td></td>
</tr>
<tr>
<td></td>
<td>p = 0.1332</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>AI acceptance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>p = 0.6293</td>
<td>No</td>
</tr>
</tbody>
</table>

When comparing 55+ (n=14) and other (n=99) participants in the home purchasing scenario regarding reliability (p (adjusted for ties) = 0.5033) and acceptance (p (adjusted for ties) = 0.8657), no significant difference was found. When analyzing 55+ (n=14) compared to other (n=96) participants in the restaurant selection sce-
nario regarding reliability (p (adjusted for ties) = 0.0111) and acceptance (p (adjusted for ties) = 0.0561), the finding for reliability was found to be a significant difference (Table 9).

The results in this study partially support H6B as participants younger than 55 were found to be more likely to rely on AI compared to the 55+ age group in the restaurant selection scenario only.

Table 9. 55+ vs. other, per scenario, AI reliability and acceptance

<table>
<thead>
<tr>
<th>Tested Variables</th>
<th>p (adjusted for ties)</th>
<th>Significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>55+ vs. other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Home purchasing</td>
<td>AI reliability</td>
<td>p = 0.5033</td>
</tr>
<tr>
<td></td>
<td>AI acceptance</td>
<td>p = 0.8657</td>
</tr>
<tr>
<td>Restaurant selection</td>
<td>AI reliability</td>
<td>p = 0.0111</td>
</tr>
<tr>
<td></td>
<td>AI acceptance</td>
<td>p = 0.0561</td>
</tr>
</tbody>
</table>

5.8 Within scenario reliability and acceptance

Besides the six hypotheses, the reliability and acceptance within each scenario was analyzed to further explore the questions to study (see chapter 1.1).

A Wilcoxon signed-rank test was used to determine if there was a statistically significant median difference between AI and human decision support regarding reliability and acceptance within each scenario. In the home purchasing scenario (n=113) 65% of participants thought the AI was reliable and 41% were willing to accept its recommendation. Regarding human decision support 53% thought the human was reliable and 39% were willing to accept the human recommendation (Table 10 and Table 11). When analyzing if participants thought the AI was more reliable than a human with p = 0.138 (n for test = 59) and if more participants were willing to accept the AI recommendation than the human recommendation with p = 0.990 in the high-risk scenario (n for test = 55), the findings in the home purchasing scenario were that no significant difference was found (Table 12).

In the restaurant selection scenario (n=110) 81% of participants thought the AI was reliable and 84% were willing to accept its recommendation. Regarding human decision support 68% thought the human was reliable and 71% were willing to accept the human recommendation (Table 10 and Table 11). Overall more participants thought the AI was reliable than the human with p = 0.015 (n for test = 54) and more participants were willing to accept the AI recommendation than the human recommendation with p = 0.033 (n for test = 56) in the restaurant selection scenario. The findings in the restaurant selection scenario were found to be a significant difference (Table 12).
Table 10. Per scenario participants grouped “Likely” to rely on AI or human recommendation

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Participant grouping</th>
<th>Reliability of AI</th>
<th>Reliability of human</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home purchasing</td>
<td>Likely</td>
<td>65%</td>
<td>53%</td>
<td>113</td>
</tr>
<tr>
<td>Restaurant selection</td>
<td>Likely</td>
<td>81%</td>
<td>68%</td>
<td>110</td>
</tr>
</tbody>
</table>

Table 11. Per scenario participants grouped “Likely” to accept AI or human recommendation

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Participant grouping</th>
<th>Acceptance of AI</th>
<th>Acceptance of human</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home purchasing</td>
<td>Likely</td>
<td>41%</td>
<td>39%</td>
<td>113</td>
</tr>
<tr>
<td>Restaurant selection</td>
<td>Likely</td>
<td>84%</td>
<td>71%</td>
<td>110</td>
</tr>
</tbody>
</table>

Table 12. Within scenario AI vs. human recommendation reliability and acceptance for participants overall

<table>
<thead>
<tr>
<th>Tested Variables</th>
<th>p</th>
<th>Significant</th>
<th>n for test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home purchasing</td>
<td>AI vs. human reliability</td>
<td>p = 0.138</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>AI vs. human acceptance</td>
<td>p = 0.990</td>
<td>No</td>
</tr>
<tr>
<td>Restaurant selection</td>
<td>AI vs. human reliability</td>
<td>p = 0.015</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>AI vs. human acceptance</td>
<td>p = 0.033</td>
<td>Yes</td>
</tr>
</tbody>
</table>

5.9 Between scenario reliability and acceptance

In addition to the six hypotheses, the reliability and acceptance between each scenario was analyzed to further explore the questions to study (see chapter 1.1). Furthermore, this analysis investigates how the risk level in the different scenarios (as opposed to perceived risk, see 5.6 above) impacts the reliability and acceptance of AI and human decision support.

A Wilcoxon signed-rank test was used to determine if there was a statistically significant median difference between AI and human decision support regarding reliability and acceptance between scenarios. In the home purchasing scenario (n=113) 65% thought the AI was reliable compared to 81% in the restaurant selection scenario (n=110) and 41% were willing to accept the AI recommendation in the home purchasing scenario compared to 84% in the restaurant selection scenario (Table 10 and Table 11). Overall more participants were likely to rely on the AI in the restaurant selection scenario compared to the home purchasing scenario with p < 0.001 (n for test = 60) and more than twice as many would be willing to accept the AI recom-
mendation in the restaurant selection scenario compared to the home purchasing scenario with \( p < 0.001 \) (n for test = 71). The findings for AI reliability and acceptance between scenarios were found to be a significant difference (Table 13).

In the home purchasing scenario 53% thought the human was reliable compared to 68% in the restaurant selection scenario and 39% were willing to accept the AI recommendation in the home purchasing scenario compared to 71% in the restaurant selection scenario (Table 10 and Table 11). Overall more participants were likely to rely on the human in the restaurant selection scenario compared to the home purchasing scenario with \( p = 0.001 \) (n for test = 69) and close to twice as many would be willing to accept the AI recommendation in the restaurant selection scenario compared to the home purchasing scenario with \( p < 0.001 \) (n for test = 71). The findings for human reliability and acceptance between scenarios were found to be a significant difference (Table 13).

These findings support the idea that people are highly willing to rely on and accept decision support in low risk situations compared to high risk situations.

Table 13. Between scenario reliability and acceptance, per scenario (high risk vs. low risk) for participants overall

<table>
<thead>
<tr>
<th>Tested Variables</th>
<th>( p )</th>
<th>Significant</th>
<th>n for test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Home purchasing vs. restaurant selection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI reliability</td>
<td>( p &lt; 0.001 )</td>
<td>Yes</td>
<td>60</td>
</tr>
<tr>
<td>AI acceptance</td>
<td>( p &lt; 0.001 )</td>
<td>Yes</td>
<td>71</td>
</tr>
<tr>
<td>Human reliability</td>
<td>( p = 0.001 )</td>
<td>Yes</td>
<td>69</td>
</tr>
<tr>
<td>Human acceptance</td>
<td>( p &lt; 0.001 )</td>
<td>Yes</td>
<td>71</td>
</tr>
</tbody>
</table>
6 Discussion

The purpose of this chapter is to discuss the findings presented in the analysis chapter. The hypotheses are revisited, the implications of the findings are presented and perspectives on the chosen method are discussed.

“How can people express "their" will if they do no have any will or conviction of their own, if they are alienated automatons, whose tastes, opinions and preferences are manipulated by the big conditioning machines?” (Paulsen, 2015:30 on "robot-ism")

6.1 Risk aversion and negative attitudes to AI

Aroyo et al. (2018:3702, 3704 and 3707) proposed a hypothesis in their study that all participants would rely on the robot during the game experiment, but risk averse people would not accept the robot’s proposal to gamble. They did this after evaluating the risk tolerance of participants. The hypothesis was partially rejected by Aroyo’s et al. (2018) findings, as people only relied on the robot when the task became more difficult, and risk averse people did accept the robot’s proposal. This is partially aligned with the findings of this study, in which the hypothesis (H1) that risk averse participants would be less likely to rely on or accept an AI recommendation, as opposed to risk seeking participants, was not supported. In fact, only 16% of participants were identified as risk averse overall which made it difficult to test the hypothesis.

Aroyo et al. (2018:3702, 3704) formulated another hypothesis in their study that people with a negative attitude towards robots would be less likely to share information. The hypothesis was rejected. This perspective could not be explored in this study as the hypothesis H2 couldn’t be tested due to lack of data.

Another result from this study is that the number of participants with negative attitudes to AI decision support was identical to the number negative to human decision support. Liang and Lee (2017:382-383) found that roughly 25% of respondents to their survey indicated fearfulness towards AI, which they relate to negative AI experiences (Liang and Lee, 2017:379). The participants of this author’s study indicate a much lower negative attitude towards AI compared to Liang and Lee (2017). This may be explained due to differences in the respective studies regarding nationality, age and gender, but primarily due to different areas of study, FARAI vs. decision-making support.

Nomura’s et al. (2006:147) results suggested that negative attitudes towards robots have a potential impact on human behavior, i.e. avoidance. A positive willingness to rely on or accept AI recommendations, as found in this study, could be interpreted
as the opposite of avoidance. Therefore, this author’s study’s findings were not aligned with Nomura et al. (2006).

Brell et al. (2018:9) found that increased experience with a technology does seem to have an impact from a risk perspective, resulting in a decrease in the overall perception of danger. This doesn’t explain why in this author’s study, so few participants had a negative attitude to AI decision support, as only 44% of the participants in this study have stated previous experience with AI assisted decision making. Additionally, the AI was described as a self-service real estate website in the home purchasing / high risk scenario and a self-service restaurant recommendation website in the restaurant selection / low risk scenario. The familiarity with website technologies and not AI per se, may have affected the participant’s willingness to rely on and accept AI recommendations. This may additionally help to explain when participants overall were more likely to rely on and accept AI recommendations, especially in the restaurant selection scenario.

Broughham and Haar (2018:245) found an overall low level of awareness regarding the impact of AI and similar technologies on the job market (STARA - Smart Technology, Artificial Intelligence, Robotics, and Algorithms). They found that awareness of STARA correlated with negative factors such as depression and cynicism. It could reasonably be expected that people fearful of losing their jobs to AI would have a negative attitude towards the technology. Since this study found participants leaning towards neutral and positive attitudes of AI that were currently employed, it could have been useful to additionally study their awareness of STARA.

### 6.2 Perceived risk and AI reliability

The hypotheses H3A and H3B looked at AI reliability in relation to perceived risk. Not only was the hypothesis rejected but additionally H3B was partially refuted since more participants found AI to be reliable compared to the humans in the restaurant scenario. Participant’s likelihood of relying on AI support in relation to the evaluated impact of the situation was the focus of hypotheses H4A and H4B. No support was found for H4A and H4B was partially refuted since more participants found AI to be reliable compared to the humans in the restaurant scenario. Perceived risk and impact in relation to reliability seem to be novel perspectives, not considered by other researchers.

Overall the participants found purchasing a house significantly riskier than selecting a restaurant and were more likely to rely on AI than a human when choosing where to have a dinner. The preference for AI in such situations may be explained by the fact that people may be unaware of the actual ability of AI to provide reliable recommendations in these types of situations and therefore incorrectly estimate AI reliability.
6.3 Acceptance of AI recommendations

Hypothesis H5 looked at participant’s willingness to accept a human recommendation rather than one made by an AI. Dietvorst et al. (2014:8) findings indicated that participants were more likely to be confident in human vs. AI ability in relation to forecasting, specifically when errors were made by the decision-making support. Lee (2018:9-10) found that human and algorithmic decision-making support could be equally trusted for mechanical skills but that humans were more trusted for human skills. Madhavan and Wiegmann's (2007:781-782) findings indicated that participants were willing to trust expert humans that made mistakes but not expert systems that were sometimes incorrect. Aroyo et al. (2018:3706) found that robots are capable of building trust, and influence people to share personal information and conform to suggestions. The results didn’t find acceptance for hypothesis H5 and partially refuted it, indicating that participants were potentially equally likely to accept AI and human in the home purchasing scenario and finding that more participants were willing to accept the AI compared to the human in the restaurant selection scenario. This result is not aligned with the findings of Dietvorst et al. (2014) and some of the findings of Madhavan and Wiegmann (2007). The findings seem to be consistent with Aroyo et al. (2018) in that participants are willing to place their trust in AIs by accepting their recommendations. Madhavan and Wiegmann (2007:775-776) found that for novice level decision-support the automation would be trusted more, which may be analogous with the low-risk, restaurant scenario.

6.4 Situational risk, reliability and acceptance

In the restaurant selection scenario, which was characterized as low risk by the majority, participants thought the AI was more reliable than the human and more participants were willing to accept the AI recommendation compared to the human recommendation. Participants were presumably more familiar with human recommendations than AI recommendations, yet they preferred AI over humans as a decision aid.

For utilizing human decision-making support, most participants only preferred doing so in the low-risk, restaurant selection scenario and not when purchasing a home. The analysis for the results of H5 indicated that participants were potentially equally likely to accept AI and human decision support in the home purchasing scenario and finding that more participants were willing to accept the AI compared to the human in the restaurant selection scenario. Overall this seems to suggest a positive attitude towards AI supported decision-making in scenarios where the perceived situational risk is low.
Lamanna and Lauren (2018:908) explored using AI as decision support in life or death situations where the subject may be incapable of making their own decisions and where an AI might give a recommendation differing from that of a doctor. Life and death situations are likely characterized by high perceived risk and high impact. This author’s study seems to confirm that people may be unwilling to rely on or accept AI recommendations in high risk situations. Therefore, it may be the nature of the situation that enables or limits reliance on decision support, whether AI or human, rather than the capabilities of or trust in the decision support itself.

Jarrahi (2018:6) found that for humans who can leverage superior creativity, imagination and intuition, decision-making will probably remain a comparative advantage vs. AI. In that sense it is quite likely that a concierge (hotel reception worker) might characterize their work in guest relations as requiring high levels of creativity, imagination and intuition in order to create fulfilling customer experiences. The work of a concierge may in fact require those soft skills, but this study seems to indicate that for decision situations characterized by low risk, such skills may add little value and fail to provide a comparative advantage for humans. Instead people may prefer to rely on and accept the recommendations of an AI. Further, this seems inconsistent with the findings of Lee (2018:14) that people perceive decisions made by algorithms as less trustworthy and less fair for tasks that people think require skills unique to humans. Human or mechanical skills may not be the deciding factor as to which decision support to rely on; rather the characteristics of the decision situation may be the key.

6.5 Gender and age perspectives

Hypothesis H6 explored the perspectives of gender and age in relation to AI reliability and acceptance. H6A stated that women would be less likely to rely on or accept AI and H6B that older age groups (55+) would be less likely to rely on or accept an AI recommendation. H6A was not supported by the data. H6B was partially supported as young participants were found to be more likely to rely on AI compared to the 55+ age group in the restaurant selection scenario only.

Liang and Lee (2017:382-383) found that woman and older people were more likely to be fearful regarding FARAI (Fear of Autonomous Robots and Artificial Intelligence). Implicitly people who exhibit FARAI would be unlikely to rely on or accept AI recommendations. Nomura et al. (2006:147) in contrast found that women had a less negative attitude to robots compared to men. This study (using Liang and Lee (2017) to inspire H6A) doesn’t contribute to confirming Liang and Lee’s (2017) or Nomura’s et al. (2006) findings regarding gender but indicate some alignment regarding age.
The findings of Broughham and Haar (2018:245) indicate that STARA (Smart Technology, Artificial Intelligence, Robotics, and Algorithms) awareness in relation to job elimination positively correlated with age among other factors. A fear of having one's job replaced by AI implies a reluctance to rely on or accept AI recommendations. This study found indications that age may indeed be a factor affecting reliance on and acceptance of AI in some situations.

6.6 Perspectives on chosen method

The survey participants were primarily residing in Sweden. Other studies have for example focused on populations in the US or Japan. This limits the ability to generalize the findings and comparisons between studies may have some limitations. Had the order of questions been randomized or the participants requested to take the survey on multiple occasions over time a different result may have been found. The survey questions and answers were inspired by previous literature. Each question was only asked in one way and answers were consistently listed from low to high values, i.e. 0-100, very unlikely to very likely, etc. Asking the same question in different ways and randomizing the questions and answers may have yielded different results, i.e. counter-balancing the questions may have minimized the influence of potential confounding variables. The author was aware of this possibility but couldn’t address it due to limitations in the online survey tool utilized.

The types of data collected limited the statistical tools applicable for analysis. Since the data was discrete, ordinal, limited in range and consisting of relatively small sample sizes, two nonparametric statistical tests were chosen for analysis.

Since this study is grounded in decision theory focusing on perspectives related to varying degrees of perceived risk in the decision situation this author chose definitions of artificial intelligence grouped as thinking human (see chapter 2.1), relating to the creation of “machines with minds” and decision-making. Other definitions are less suitable as they focus on studying computation, intelligent agents or machines that perform human-like functions. There are currently many definitions of AI and little agreement between researchers on what is the suitable in each situation. From a survey participant perspective, it is recognized that AI may mean many things, including computers, algorithms, robots, machine-learning, etc. It is recognized that this freedom of interpretation may limit the validity of the results but wasn’t practically possible to ensure that everyone had the same view.

It is challenging to identify a perspective on risk (see chapter 2.2.3) that is uniformly shared by researchers and the public, i.e. survey participants (for example Drottz-Sjöberg and Sjöberg, 2003:316). Consequently, the notion of risk used in this study is based on perceived risk and impact from an investigator’s perspective, aligned with the perceived risk and evaluated impact of different scenarios by the partici-
pants. A potential misalignment of perceived risk may limit the validity of the results, for example if the survey participants didn’t agree with the description of different scenarios as high or low risk. Some of the questions in the survey were intended to mitigate this potential problem.

Kahneman and Tversky (1979:271) showed that people make inconsistent decisions when the same choice is presented in different ways. This study may not have been optimally designed as the potential gains (selecting a good restaurant) or losses (risk of purchasing the wrong home) were only presented in one way. Had the scenarios been presented differently the participants may have had other preferences regarding their willingness to rely on and accept AI recommendations. More than two scenarios with different risk levels could have been presented as well as multiple scenarios with similar risk levels. Additionally, the order the scenarios were presented in could have been randomized for different participants.
7 Conclusion

The purpose of this chapter is to present the overall answer to the investigated question and the outcomes of the tested hypotheses. The contributions to theory are described including suggestions for future research.

7.1 Study outcomes

We are

“living within nineteenth-century social systems, based on seventeenth-century ideologies . . . we must terraform not only Mars, but ourselves.” (Kim Stanley Robinson via Sargeant, 2018)

The purpose of this study was to explore the attitudes to AI decision-making support in risky contexts. The question this paper posed was: How are people’s attitudes to artificial intelligence vs. human decision-making support affected by the level of perceived risk in the decision situation? The question was explored in detail through formulating and testing six hypotheses. The results for the hypotheses are described in Table 14 below.

Table 14. Results of testing the six hypotheses

<table>
<thead>
<tr>
<th>Index</th>
<th>Hypothesis</th>
<th>Accepted</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>Risk averse participants will be less likely to rely on or accept an AI recommendation</td>
<td>No</td>
</tr>
<tr>
<td>H2</td>
<td>Participants with a negative attitude to AI will be less likely to rely on or accept an AI recommendation</td>
<td>No</td>
</tr>
<tr>
<td>H3A</td>
<td>Participants will be more likely to rely on human rather than AI support in a decision situation characterized by high perceived risk</td>
<td>No</td>
</tr>
<tr>
<td>H3B</td>
<td>Participants will be equally likely to rely on human and AI support in a decision situation characterized by low perceived risk</td>
<td>Partially refuted</td>
</tr>
<tr>
<td>H4A</td>
<td>Participants will be more likely to rely on human rather than AI support in a decision situation characterized by high impact</td>
<td>No</td>
</tr>
<tr>
<td>H4B</td>
<td>Participants will be equally likely to rely on human and AI support in a decision situation characterized by low impact</td>
<td>Partially refuted</td>
</tr>
<tr>
<td>H5</td>
<td>Participants will overall be more likely to accept a human than an AI recommendation</td>
<td>Partially refuted</td>
</tr>
<tr>
<td>H6A</td>
<td>Women will be less likely to rely on or accept an AI recommendation</td>
<td>No</td>
</tr>
<tr>
<td>H6B</td>
<td>Older age groups (55+) will be less likely to rely on or accept an AI recommendation</td>
<td>Partially supported</td>
</tr>
</tbody>
</table>

At an overall level and as an answer to the study question, people do seem to be affected by the perceived level of risk in the decision situation. The results seem to suggest a positive attitude towards relying on and accepting AI decision-making in scenarios where the perceived situational risk is low, such as selecting a restaurant. For higher risk situations the results seem to indicate a reluctance to rely on or ac-
cept either AI or human advice. It is not therefore a given that people will appreciate the potential unique value that a human advisor could provide as opposed to an AI.

The implication of this study is that AI may currently be best applied to situations characterized by perceived low risk if the intention is to convince people to rely on and accept AI recommendations, and in the future if AI becomes autonomous, to accept decisions.

AI has an increasing social impact (Russell and Norvig, 2010:1034-1040). Biases in the data used to train an AI may result in discrimination. People may lose their jobs or not be hired in the first place. There is a risk that AI may result in people making unethical decisions. By conducting this investigation, it is the hope of this author that a contribution has been made to better understanding human and AI interactions, hopefully resulting in a decreased likelihood of discrimination and an increased likelihood of ethical decisions.

### 7.2 Future research

For future research there are several topics of interest. One would be to retest the hypothesis in a similar manner to investigate if the result is the same. Another would be to expand the number and scope of the risky scenarios, as well as exploring risk in terms of both absolute and relative losses and gains. The participants in the study have a quite narrow age range, were mostly men and primarily resided in Sweden. Studying other populations might yield other interesting perspectives. Additionally, it might be useful to conduct the same survey with the same participants over time to evaluate how and if their attitudes change as AI becomes more prevalent in society.
References


Appendix A

In this appendix is the survey that was published and made available to the participants in the online tool.

Note: Text between < and > was NOT visible to survey participants

Welcome! <Page 1>

The purpose of this survey is to investigate your attitudes to decision-making support under risk where your decision-making is supported by other people or by artificial intelligence (AI).

You will be presented with two different scenarios with varying degrees of risk and you will be asked to evaluate your reliance on humans and AI (artificial intelligence) in supporting you to decide in each case.

You are requested to answer: 6 risk and decision support questions, 12 scenario questions and 6 additional questions (24 questions total). The expected time to complete the survey is ~10 minutes.

If you have any questions or concerns please contact: pernordahl13@gmail.com

Note: The survey is anonymous. All data stored in this online tool will be deleted after 1 month. Data may be retained offline by the researcher as needed to complete the research.

Risk & Decision Support Questions <Page 2>

1. When you think of the word "risk" which of the following words comes to mind first?
   Loss / Uncertainty / Opportunity / Thrill

2. You are on a TV game show and can choose one of the following. Which would you take?
   $1,000 in cash / A 50% chance at winning $5,000 / A 25% chance at winning $10,000 / A 5% chance at winning $100,000

3. In general, how would your best friend describe you as a risk taker?
   A real risk avoider / Cautious / Willing to take risks after completing adequate research / A real gambler

4. You have just finished saving for a "once-in-a-lifetime" vacation. Three weeks before you plan to leave, you lose your job. You would:
Cancel the vacation / Take a much more modest vacation / Go as scheduled, reasoning that you need the time to prepare for a job search / Extend your vacation, because this might be your last chance to go first-class

5. What is your overall attitude to utilizing AI (artificial intelligence) to help you in your decision-making?

Very negative / Somewhat negative / Neither negative nor positive / Somewhat positive / Very positive

6. What is your overall attitude to utilizing other people to help you in your decision-making?

Very negative / Somewhat negative / Neither negative nor positive / Somewhat positive / Very positive

**Scenario A <Page 3>**

You are planning to purchase a new home valued at $500,000 USD. You have realized that you would like some support in helping you decide which home to buy. You would like the most value for money for your new home according to your own interpretation of “value”.

There are different possible outcomes for this scenario. You might choose a bad home, an average home or a good home. Use your imagination to clarify what this means for you.

7. How high is the perceived risk for you when purchasing a new home valued at $500,000 USD? <Brell at al. (2018:10)>

0% - 100%

8. How do you think this home purchase will impact your life overall?

Very small impact / Somewhat small impact / Neither large nor small impact / Somewhat large impact / Very large impact

**Decision Support A <Page 4>**

You have two alternatives to choose from to help you make your decision:

1) A self-service real estate website based on AI (artificial intelligence) from a well-known brand.

2) A human real estate agent at a well-known real estate agency.

The outcomes for the AI (artificial intelligence) decision support can be as follows: the real estate website suggests a bad home, an average home or a good home.
9. How likely do you think it is that the AI (artificial intelligence) will suggest a recommendation that you can rely on?

Very unlikely / Somewhat unlikely / Neither likely nor unlikely / Somewhat likely / Very likely

10. How likely are you to accept the home recommended by the AI (artificial intelligence)?

Very unlikely / Somewhat unlikely / Neither likely nor unlikely / Somewhat likely / Very likely

The outcomes for the human decision support can be as follows: the real estate agent suggests a bad home, an average home or a good home.

11. How likely do you think it is that the real estate agent will suggest a recommendation that you can rely on?

Very unlikely / Somewhat unlikely / Neither likely nor unlikely / Somewhat likely / Very likely

12. How likely are you to accept the home recommended by the real estate agent?

Very unlikely / Somewhat unlikely / Neither likely nor unlikely / Somewhat likely / Very likely

**Scenario B <Page 5>**

You are planning to attend a dinner valued at $50 at a restaurant. You have realized that you would like some support in helping you decide which restaurant to have dinner at. You would like the most value for money for your meal according to your own interpretation of “value”.

There are different possible outcomes for this scenario. You might have a bad dinner, an average dinner or a good dinner. Use your imagination to clarify what this means for you.

13. How high is the perceived risk for you when choosing a restaurant to have a dinner valued at $50 USD? <Brell at al. (2018:10)>

0% - 100%

14. How do you think this dinner will impact your life overall?

Very small impact / Somewhat small impact / Neither large nor small impact / Somewhat large impact / Very large impact
**Decision Support B <Page 6>**

You have two alternatives to choose from to help you make your decision:

1) A self-service restaurant recommendation website based on AI (artificial intelligence) from a well-known brand.

2) A human concierge (reception worker) at a well-known hotel.

The outcomes for the AI (artificial intelligence) decision support can be as follows: The recommendation website suggests a restaurant that results in a bad dinner, an average dinner or a good dinner.

15. How likely do you think it is that the AI (artificial intelligence) will suggest a recommendation that you can rely on?

- Very unlikely
- Somewhat unlikely
- Neither likely nor unlikely
- Somewhat likely
- Very likely

16. How likely are you to accept the dinner recommended by the AI (artificial intelligence)?

- Very unlikely
- Somewhat unlikely
- Neither likely nor unlikely
- Somewhat likely
- Very likely

The outcomes for the human decision support can be as follows: the concierge suggests a restaurant that results in a bad dinner, an average dinner or a good dinner.

17. How likely do you think it is that the concierge will suggest a recommendation that you can rely on?

- Very unlikely
- Somewhat unlikely
- Neither likely nor unlikely
- Somewhat likely
- Very likely

18. How likely are you to accept the dinner recommended by the concierge?

- Very unlikely
- Somewhat unlikely
- Neither likely nor unlikely
- Somewhat likely
- Very likely

**Additional Questions <Page 7>**

19. Age?

- 18-24
- 25-34
- 35-44
- 45-54
- 55-64
- 65+

20. What is your gender? (With gender we mean gender identity, i.e. the gender you feel yourself to be)
21. Which country are you currently in?
[Country list]

22. What is your current work status?
Student / Working (Part-time, full-time, self-employed) / Retired / Other

23. Have you ever used an AI (artificial intelligence) to assist you in real decision-making (as far as you know)?
Yes / No

24. Please share any thoughts or feelings you may have about using AI (artificial intelligence) to assist people in making decisions?
[Open-ended]

Thank you for completing this survey!