Teaming With Robots: Do Humans Judge Decisions Made by Robots Differently Than Decisions Made by Humans?

Maike C. Eversmann

s1683926

Psychology of Conflict, Risk and Safety

1st Supervisor: Prof. Dr. José H. Kerstholt

2nd Supervisor: Drs. Esther S. Kox

University of Twente

March 10, 2022

Abstract

Technology is nearly indispensable nowadays and the use of robots increased in recent years. Robots are employed in various fields to give advice and assist humans in making decisions, or to even make decisions by themselves. Since robots are often used in risky situations and their influence will further increase in the future, it is important to study human-robot interaction. In an online experiment, the present study investigates how people judge decisions made by robots compared to decisions made by a human expert. As a second goal, the study examines whether the judgment of decisions is dependent on the situation. To answer these questions, a 2 (Decision maker: human vs. robot) x 2 (Decision: biased vs. rational) between-subjects design with six different Scenarios as a within-subjects variables has been chosen. After each scenario, participants evaluated the decision of the decision maker. The results of the study showed that the decisions made by the robot were not judged differently than the decisions made by the human expert. The omission of outcomes in the scenarios as well as the inclusion of justifications has been identified as possible explanations. An additional finding of the study was that the judgment of the decision was dependent on the situation. The rational decision was evaluated more positive than the biased decision in three scenarios while no difference in the judgment between the decisions was found in the other three scenarios. A possible explanation of this result was that the bias was more apparent in the three scenarios in which the decisions were judged differently than in the scenarios in which both decisions were judged equally.

Keywords: robots, artificial intelligence, human-robot interaction, decision making

Teaming With Robots: Do Humans Judge Decisions Made by Robots Differently Than Decisions Made by Humans?

Technology increasingly gained importance in everyone's life. Devices such as smartphones and notebooks have replaced their non-technical counterparts and the internet gives access to more information than ever before. People increasingly rely on technology because it provides them with help in many situations (Shneiderman, 2007). Therefore, such devices are nearly indispensable nowadays. In recent years, the development of devices that are based on artificial intelligence (AI) increased. AI can be defined as "a computer system performing tasks that normally require human-level intelligence, like the identification of patterns in data and the generation of predictions" (Biller-Andorno et al., 2021, p. 175). Algorithms, machine and deep learning enable AI to perform such complex tasks (Biller-Andorno et al., 2021; Tanibe et al., 2017). Based on AI, robots were developed and introduced in our daily life (Tanibe et al., 2017). Robots are a form of AI and therefore, have the same technical abilities. In addition, robots can act and judge autonomously (Tanibe et al., 2017). Their presence can be either virtual or physical (Huang et al., 2021; Shinozawa et al., 2005). Although the use of robots is common already, their impact will further increase in the future (Tanibe et al., 2017).

Human decisions are rarely made alone but are often influenced by input from other people or robots. People often ask for advice before making a decision (Dalal & Bonaccio, 2010; Tzioti et al., 2013). The advice of robots can be helpful because they can make fast and accurate decisions (Bahner et al., 2008). However, failures in algorithms of robots or advice that is not in line with the preferences of the decision maker can cause suboptimal decisions. Therefore, it is important to study the interaction of humans and robots in terms of decision making, especially because robots are often used in risky situations (Skitka et al., 2000). Furthermore, it is important to explore how humans judge decision made by robots. Since robots are increasingly used in decision making, it is relevant to determine whether people value the decisions of humans more than those of robots (Huang et al., 2021; Lieberman, 2001).

The purpose of this study is to explore how decisions of robots are evaluated compared to identical decisions of humans in the same situation. Hence, the study addresses the following research question: *Do humans judge decisions made by robots differently than decisions made by humans?* Robots are employed in various fields and therefore make decisions in a multitude of different situations. Thus, the second research question is *whether the judgment of decisions depend on the situation*.

Decision Making

Human decision making is based on two thinking systems (Evans, 2008). System 1 is fast, intuitive and happens unconsciously. Additionally, it is influenced by affect and mainly based on heuristics. Heuristics simplify a complex problem by using a rule of thumb (Kahneman, 2012). In contrast, System 2 is slow, controlled and conscious. It is rule-based and describes analytic thinking. Most decisions humans make are based on System 1 thinking (Gladwell, 2007). Although intuitive decision making can be effective (Dane et al., 2012), it can also cause biased decisions (van den Bosch & Bronkhorst, 2018). One example of a heuristic leading to biased decisions is the availability heuristic (Slovic et al., 1980). How available an event is in someone's mind, affects their judgment of the event (Slovic et al., 1980). For example, if people are asked to judge the risk of a natural hazard, their judgment is influenced by how easily they can remember the occurrence of or information about this natural hazard. While most decisions are made based on System 1 thinking, System 2 thinking is needed to make analytic and well-thought decisions.

To make such analytic and well-thought decisions, people may ask someone else for advice (Tzioti et al., 2013), particularly in complex situations. Advice taking is influenced by different factors such as trust, competence, and source of advice (Dalal & Bonaccio, 2010; Tauchert & Mesbah, 2019).

Trust influences whether people take advice or reject it (Jungermann, 1999). According to Mayer et al. (1995), trust is "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" (p. 712). Trust is a multidimensional construct and is influenced by different perceived characteristics of the trustee such as ability, benevolence and integrity (Mayer et al., 1995). Ability describes the perceived competence of a person in a specific domain (Mayer et al., 1995). A high perceived ability of the advisor has a positive impact on the acceptance of the advice (Schultze et al., 2015). The concept of benevolence revolves around the egocentric motive of the trustee (Mayer et al., 1995). If the trustor perceives that the trustee has a high egocentric motive, the perceived benevolence of the trustee is low. On the contrary, if the trustee seems to act in the interest of the trustor, the perceived benevolence is high. Advice is more likely to be unbiased and in the interest of the person who asked for advice if the advisor has high benevolence (Komiak & Benbasat, 2006). Lastly, integrity refers to the trustor's acceptance of the principles of the trustee as well as the trustee's adherence to these principles (Mayer et al., 1995). The possibility that the advice will be used is higher if the integrity of the advisor is high (van Swol, 2011). The higher the ability, benevolence and integrity of the advisor, the higher the trust in the advisor. Advice is likely to be accepted if trust in the advisor is high. However, if trust is low, it is possible that the advice will be rejected (Jungermann, 1999). Trust is therefore a key factor in the acceptance of advice.

Another factor that influences advice taking is competence. For the person who asks for advice, it is important to know that the advisor has the competence to do so (Komiak & Benbasat, 2006). A competent advisor is capable to provide an advice that improves the decision (Komiak & Benbasat, 2006). The perceived competence influence trust in the advisor but it also influences whether people change their decision based on the advice given (Schultze et al., 2015). Competence is one of the factors that influence advice taking and it influences it directly but also indirectly through trust.

The source of advice also has an impact on advice taking. Advice from people perceived to be more experienced is more likely to be followed than advice from people who have the same or less experience than the decision makers themselves (Harvey & Fischer, 1997). If the source of the advice is a robot, this has shown to have an impact on the utilization of advice. While the advice of robots is valued in some situations (Tauchert & Mesbah, 2019), this is not necessarily the case in other situations (Starr et al., 2021). Thus, the source of advice is also an important factor in advice taking and can influence the decision to accept or reject the advice.

Human-Robot Interaction in Decision Making

People are not only asking other humans for advice, though, they are also increasingly relying on robots (Gonzalez Fabre et al., 2020; Rossi & Mattei, 2019). Robots are able to analyze a lot of data in a short period of time (Biller-Andorno et al., 2021). Accordingly, robots' decisions are faster and often more accurate than human decisions, especially in situations that include much information (Bahner et al., 2008; Shneiderman, 2016). However, this does not mean that people have to accept the decisions made by a robot.

Just like the interaction with humans, the interaction with robots is also influenced by trust (Bisantz & Seong, 2001; Freedy et al., 2007; Huang et al., 2021). Low levels of trust in the robot can lead to rejection of important information only because a robot has provided it. High levels of trust, however, can lead to biased decisions and poor outcomes if the information provided by robots is left unquestioned and blindly followed (Hidalgo et al., 2021). The greater the trust in robots, the higher the likelihood that people will accept the information provided by robots.

Trust in a robot is influenced by different factors. One of these factors is the perceived ability of a robot to make an accurate decision. The findings of Huang et al. (2021) showed that high-ability robots receive a higher level of trust than low-ability robots. Errors made by the robot reduced the level of trust. Furthermore, the robot's ability to communicate also plays a role in building trust (Hancock et al., 2011; Volante et al., 2018). Volante et al. (2018) and Sanders et al. (2014), for example, showed that people trust robots more if the robot can communicate with them. Additionally, Sanders et al. (2014) who examined the effects of transparency and communication on trust, discovered that robots that constantly shared information with their human team members were trusted more than robots that shared little or no information. Another characteristic influencing trust is the attitude towards robots (Volante et al., 2018). Negative attitudes towards robots can lead to negativity bias and even algorithm aversion which could affect trust in robots (Hidalgo et al., 2021). Thus, trust in robots is impacted by perceived ability, communication, and attitude towards robots.

In addition to trust, morality is also an important factor in human-robot interactions. Moral norms build an essential part of society. They guide people how they should behave and interact with each other (Malle et al., 2015). Since robots can make their own decisions and act autonomously (Tanibe et al., 2017), it is important to question their morality in human-robot interaction (Gonzalez Fabre et al., 2020; Greene et al., 2019). Previous research found that people judge robots differently than humans when they make a moral decision (Hidalgo et al., 2021). It revealed that a utilitarian choice is more expected from robots than from humans (Malle et al., 2015). A utilitarian choice means that one person is sacrificed for the well-being of many. Additionally, robots were also more strongly blamed than humans if they did not make a utilitarian choice (Malle et al., 2015). Whether robots make such a utilitarian choice depends on their algorithm. Sometimes, however, the algorithms of a robot include faults (Biller-Andorno et al., 2021) which can lead to biased decisions. These decisions have the potential to raise additional moral questions. Therefore, it is important to consider morality as a factor in human-robot interactions.

Decision Making Biases

Previous studies focused primarily on social and moral dilemmas but there are other decision biases that can be used to study human-robot interactions. This study examines a social dilemma as well as five of other decision biases, including sunk cost bias, framing bias, mental accounting, affect as heuristic, and time discounting.

A social dilemma is a situation in which one person is sacrificed for the good of many (Bonnefon et al., 2016). An example of a social dilemma is self-driving cars. They are developed to reduce accidents and therefore, make driving safer (Bonnefon et al., 2016). However, sometimes an accident cannot be avoided. Therefore, the self-driving car must include a programmed rule that saves either the driver's life or the lives of several others (Goodall, 2014). Including such a moral rule creates a social dilemma because either option means the other party can lose their life (Bonnefon et al., 2016). When faced with a social dilemma, the person making the decision must choose which life to save.

The second decision bias, the sunk cost bias, refers to the continuation of a project that is no longer rewarding (Emami et al., 2019; Haita-Falah, 2017; Strough et al., 2016). Sunk costs include all investments already made for the project such as money or time (Garland & Newport, 1991). Since sunk costs are prior investments, they should not be considered when deciding on the continuation of the project (Garland & Newport, 1991). If people include sunk costs into their decision-making process nonetheless, they subject themselves to the sunk cost bias.

In addition to sunk costs, the presentation of a problem can also led to decision bias, namely framing bias. A situation can be framed either in terms of gains or in terms of losses (Hodgkinson et al., 2002). If a problem is framed in terms of gains, the potential benefits are highlighted. Consequently, people tend to prefer risk-averse alternatives and reject the risky option (Bless et al., 1998; Emami et al., 2019). On the contrary, if the same problem is framed in terms of losses and the losses are emphasized, people tend to prefer risk-taking alternatives (Bless et al., 1998; Emami et al., 2019). Hence, a person's decision is influenced by the way the problem is formulated (Haita-Falah, 2017; Tan & Yates, 1995).

Another bias that influences the decision making of people is mental accounting (Emami et al., 2011; Emami et al., 2019; Kahneman, 2003). Mental accounting addresses the bias that judgments are relative rather than absolute. In the example of cost savings, the mental accounting effect contains that "the same absolute saving on an item appears to be more attractive the higher its value *relative* to the item's original price" (Duxbury et al., 2005, p. 568). If people can save the same absolute amount of money for a cheap item or an expensive item and have to invest exactly the same amount of time to do so, they are more inclined to save money for the cheap item than for the expensive item. People choose the cheap item because the relative savings are higher than those of the expensive item (Duxbury et al., 2005).

The fifth bias, affect as heuristic, concerns emotions which cause bias in decision making (Loewenstein & Lerner, 2003). An illustration of this heuristic is how lives are valued and how decisions are made about saving lives (Västfjäll et al., 2014). Normally, the willingness to save lives should be higher when the number of people threatened increases. It has been observed, however, that the compassion rate is highest if only one person's life is threatened and decreases with each increase in the number of people at risk (Kogut & Ritov, 2005). An example of affect as heuristic bias is the identifiable victim effect with respect to donations. This phenomenon means that people are more willing to donate if there is only one identifiable person in need than if there are multiple victims (Jenni & Loewenstein, 1997). People's emotions lead to the decision to donate to only one person instead of helping many people.

Lastly, time discounting refers to comparing the values of rewards that are available in the present or in the future (Ranyard et al., 2006). Future rewards are discounted because of the delay in time. This means that the value of future rewards is discounted relatively for anything that might reduce the expected reward such as changes in taste or inflation (Frederick et al., 2002; McClure et al., 2007). Time discounting cause people to prefer a lower immediate reward and reject a higher future reward which is called time preference (Frederick et al., 2002).

The Present Study

The present study aimed to examine whether humans judge the decisions made by a robot differently than the same decisions made by a human. Moreover, the study investigated whether the judgment changes in different decision tasks.

The design of the study was inspired by the study of Hidalgo et al. (2021). The present study extends the original study by referring to different decision biases in the scenarios, rather than just social dilemma situations. In addition, the present study contained a justification of the decision by the decision maker. Outcomes or consequences of the decision which were included in the study of Hidalgo et al. (2021), however, were not incorporated.

To answer the research questions, the judgment of decisions was measured in different scenarios. The decision maker was either a human or a robot. Based on the research questions and the theoretical framework, the following hypotheses have been derived:

H1: Decisions made by a robot are judged more negatively than decisions made by a human.

H2: This effect is stronger for biased decisions than for rational decisions.

H3: This effect is stronger if their decisions have a negative impact on people than if the decisions would not have a negative impact on people.

Methods

Design

The study had a mixed experimental design with a 2 (Decision maker: human vs. robot) x 2 (Decision: biased vs. rational) between-subjects design and Scenario as a within-subjects variable. The experiment took place online. Each participant was presented with six different Scenarios in which the decision maker has been consulted to make a decision. The Scenarios were displayed in a fixed order.

The dependent variables were trust in the decision maker and their perceived appropriateness for the task, as well as the moral acceptance, sensibility, and agreement with the decision. Attitude towards AI was included as a control variable to analyze whether the participants' attitude had an impact on their judgment of the decision and the decision maker.

Participants

Participants were recruited through convenience sampling. An invitation message with information about the study and a link to the online experiment was distributed among social contacts as well as on Social Networking Sites such as Facebook and Instagram. Invited were people who were at least 18 years old. Overall, 134 people participated in the study. Exclusion criteria were denial of the informed consent (n = 2) and withdrawal from the study (n = 48). A total of 50 people were excluded from the study, resulting in a sample of 84 participants. All participants were randomly divided into four groups. The Human-Biased condition included 22 participants, 21 participants were in the Human-Rational condition, whereas 20 participants were in the Robot-Biased condition, and the Robot-Rational condition included 21 participants.

The mean age of participants was 32.4 years (SD = 14.0) with a range from 19 years to 87 years. The sample included 37 males (44.0%), 46 females (54.8%) and one non-binary person (1.2%). Dutch was the nationality of five participants (6.0%), whereas 64 participants were German (76.2%) and 15 participants indicated to have another nationality than Dutch or German (17.9%).

Materials

The study included six different scenarios which were based on decision making biases. Each scenario was formulated independently and had no reference to one of the other scenarios. After the description, two decision options were presented, followed by the decision of the decision maker, including the justification of the decision. Each scenario contained a decision option that was biased according to the underlying decision bias, and a rational decision option.

Identical decisions and justifications were included for the human and the robot decision maker. All scenarios, including decision options and justifications, are presented in Appendix A.

The first scenario was based on the *sunk cost bias*. A project which has been worked on for quite some time and is almost completed is not profitable according to current customer surveys. A decision has to be made whether to continue the project or to terminate it.

Scenario 2 was based on the *framing bias*. It involves a situation in which a village is about to burn down. The decision makers can decide to evacuate the village and save 50% of the inhabitants, or they can fight the fire and save all inhabitants with a probability of 50%.

A *social dilemma* was presented in the third scenario. In this scenario, a self-driving car had to be programmed. In a situation where three pedestrians suddenly cross the street, the self-driving car can either crash into a tree and risk the driver's life or run over the pedestrians to save the driver's life. The decision maker must decide which option to choose.

In scenario 4 which was based on *mental accounting*, the person wants to buy tickets. In the first ticket shop, the tickets cost 100€. With a walk of ten minutes to the second store, the same tickets could be bought for 95€. In this scenario, the decision maker must decide where to buy the tickets. To do this, the decision maker has to evaluate whether to invest the time to save the 5€ or to pay the full price.

Based on the *affect as heuristic* bias, the fifth scenario described a situation in which 100€ could be donated to one of two nonprofit organizations. The first organization was described as helping several people in Africa whereas the second organization stated that the donation would support a boy named Chad. The decision maker must decide to which organization will receive the donation.

The last scenario made use of *time discounting*. A choice could be made between two bonuses. The bonus can be paid either immediately or at the end of the year. The bonus paid at the end of the year is 50€ higher than the immediate bonus. In this scenario, the decision had to be made which bonus to choose.

After each scenario nine questions were posed that measured the dependent variables of the study for the scenario. Examples were *The decision is morally acceptable*, *The other decision option would have been the better alternative*, and *The decision maker should be replaced by a human/robot*. All items were measured on a 5-point Likert scale from 1, strongly disagree, to 5, strongly agree. At the end of the study, participants were asked to answer nine questions about their attitude towards AI. *I am worried when I think of AI* and *In my opinion AI solves problems* were examples. Eight of the nine items were measured on a 5-point Likert scale from 1, strongly disagree, to 5, strongly agree. The exception was the item *I heard about artificial intelligence (AI)* which was measured on a 5-point Likert scale from 1, not at all, to 5, a lot. The scale measuring Attitude towards AI was reliable, alpha = .79. Appendix B includes all items used to measure the demographic, dependent, and control variables.

Procedure

After participants received the invitation to participate in the study and entered the online experiment via the provided link, they were presented with the informed consent. They were informed that all data will be anonymized, that the study is in accordance with the ethical norms, and that they can withdraw from the study at any time. Participants were also informed that the purpose of the study was to determine how people evaluate different decisions in various situations. The actual aim of the study, to investigate the judgment of decisions made by robots compared to decisions made by other humans, was withheld from participants to avoid possible biases. See Appendix C for the Informed Consent of the study. If participants agreed to the informed consent, they were randomly assigned to one of four conditions. Participants who denied informed consent were redirected to the end page of the study.

Next, participants were given a brief introduction explaining that a human expert / robot was consulted in various situations to make a decision. Some information about the decision maker was given. Appendix D includes the introduction for both the human and the robot condition. Afterwards, participants were asked to answer some questions on demographics.

Subsequently, participants received the scenarios. Each scenario included a description of the scenario, two decision options, and the decision including the decision maker's justification. After each scenario, participants were asked to answer some questions about the decision and the decision maker to measure the dependent variables. After the last scenario, participants were presented with a survey containing nine items measuring Attitude towards AI.

Lastly, participants were debriefed and thanked for their participation. The debriefing included an email address for potential questions and further information. See Appendix C for the debriefing of the study.

Analysis

The data analysis was performed in the statistics program SPSS. Before the analysis, some of the items had to be recoded because they were formulated in such a way that their scales were reversed. The recoded items are indicated in Appendix B. The nationality of the participants was asked using a free textbox. For the analysis, the nationality of participants was categorized into German, Dutch and nationalities other than German or Dutch.

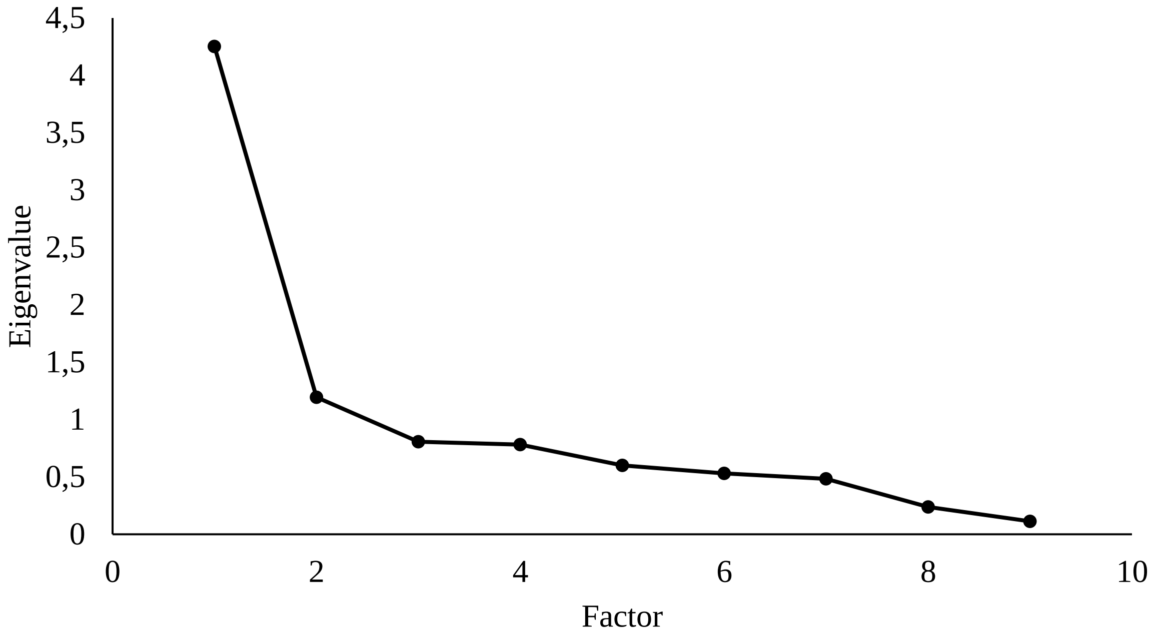
Results

Factor Analysis

To determine the number of factors of the dependent variables and whether the items measured the intended variables or whether some items can be combined into one variable, an exploratory factor analysis with a Principal Component extraction and a Varimax rotation was performed for Scenario 1. Scenario 1 was used as an example for all scenarios. The scree plot which is illustrated in Figure 1 was used to determine the number of factors. According to the Elbow Criterion which identifies the dimensionality based on the position of the bend, the dependent variables had a dimensionality of two factors.

Figure 1

Scree plot of the Factor Analysis of the Dependent Variables in Scenario 1



The rotated factor matrix was used to identify the factor loading of each item. The cut-off value for the correlations was .30 with items that showed a correlation higher than .30 loaded on the corresponding factor. The results of the factor analysis showed that all nine questions loaded either on factor 1 or on factor 2. Appendix E includes the rotated factor matrix with all factor loadings of the nine items that are above the cut-off value of .30. Items 1 to 7 loaded on factor 1 whereas items 8 and 9 loaded on factor 2. Item 8, *The decision maker should be replaced by a human / robot*, showed a negative correlation with factor 2. All other items had a positive correlation with the corresponding factor.

Based on the results of the factor analysis, new variables were created by calculating the mean of all items that were correlated to the same factor. The first seven items were summarized into the variable *Judgment* which represents the judgment of the decision and the decision maker. Item 8 and 9 were summarized into the variable *Appropriateness* which represents the perceived appropriateness of the decision maker to make this decision. A low score in this variable would mean that the decision maker is not perceived as appropriate to make the decision and should better be replaced by another decision maker. The variables were adopted for all scenarios.

Correlations

Correlations were calculated to analyze possible intercorrelations between demographics, Attitude towards AI, and the dependent variables Judgment and (perceived) Appropriateness. For the calculation of the correlations, the means of Judgment and (perceived) Appropriateness across scenarios were used. No significant correlations above the threshold value of .30 were shown. Nevertheless, a weak correlation was found between age and Attitude towards AI, $r(82) = -.25$, $p = .024$. Nationality and Attitude towards AI also showed a weak correlation, $r(82) = .24$, $p = .025$. Furthermore, a weak positive correlation was found between Judgment and (perceived) Appropriateness, $r(82) = .26$, $p = .019$. No other significant correlations were found. See Appendix F for the correlation matrix.

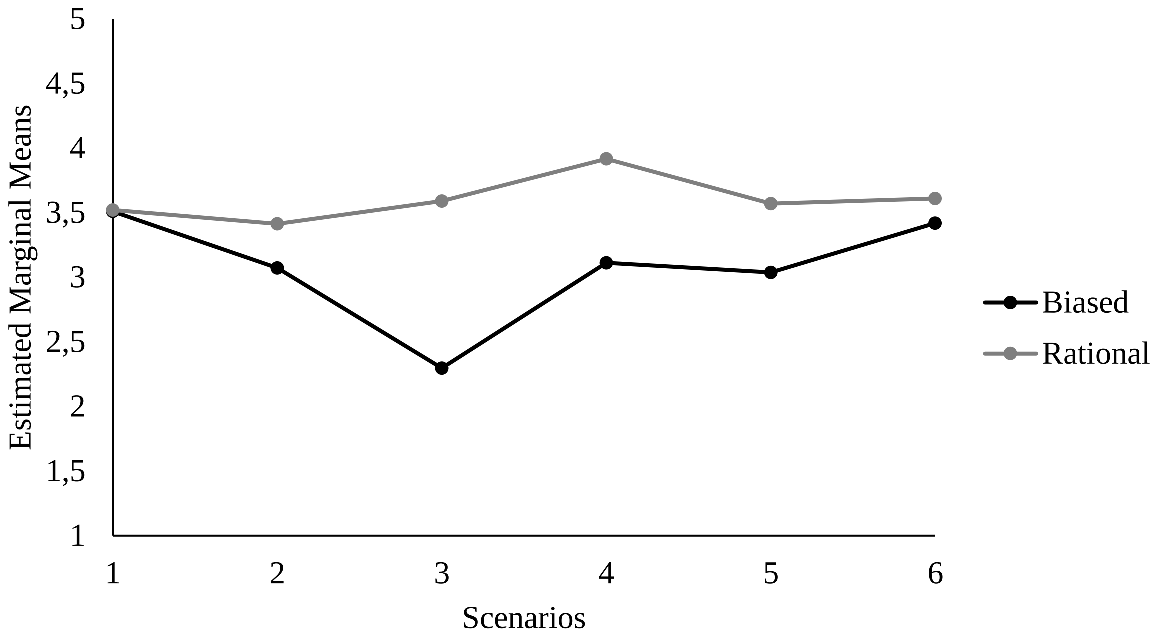
Judgment

To analyze the effects of the independent variables on the dependent variable Judgment, and to detect differences between groups, a repeated ANOVA analysis was performed. Scenarios 1 to 6 were included as levels in the analysis. Between-subjects factors were the Decision (biased vs. rational) and the Decision maker (human vs. robot). Attitude towards AI was included as covariate. The Box's test of equality of covariance matrices was not significant, $p = .139$, indicating that there is no evidence for violation of homogeneity of covariance matrices. The multivariate tests showed only a significant interaction between Scenarios and Decision, Wilks' lambda = .62, $F(5, 75) = 9.23$, $p < .001$. All other main effects and interactions had no significant effect, $p > .05$.

To test whether the differences between the Decisions within Scenarios are significant, an Interaction contrasts analysis has been performed. A one-factorial ANOVA contrasts analysis for the influence of Decision on Judgment was conducted. The biased decision was allocated to coefficient -1 whereas the rational decision was allocated to coefficient 1. The analysis showed a significant difference in Judgment between the Decisions for Scenario 3, *t*(82) = 7.48, *p* < .001. The analysis also revealed significant results for Scenario 4, *t*(82) = 5.74, *p* < .001, as well as for Scenario 5, *t*(82) = 3.41, *p* < .001 (see Figure 2). All three scenarios had positive t-values indicating that the difference is significant according to the assumption that the rational decision has a positive coefficient compared to the biased decision. The differences in Judgment were not significant for the remaining Scenarios, p > .05.

Figure 2

Profile plot of judgment for the interaction between Scenarios and Decision on Judgment



(Perceived) Appropriateness

A repeated ANOVA analysis was also performed to analyze the effects of the independent variables on Appropriateness and to detect differences between groups regarding their (perceived) Appropriateness of the Decision maker. The same levels, between-subjects factors and covariate were used as in the repeated ANOVA analysis for Judgment. According to the Box's test of equality of covariance matrices, there is evidence that there is a violation of homogeneity of covariance matrices, *p* = .025. Pituch and Stevens (2016) suggested that the results of the multivariate tests are fairly robust if the ratio of the largest n to the smallest n is smaller than 1.5. The ratio of the largest n to the smallest n in this study was 22/20 = 1.1. Therefore, it can be assumed that the results of the multivariate tests are fairly robust even if the Box's test of equality was significant. The multivariate tests revealed no significant main effect for any main or interaction effect, *p* > .05.

Discussion

The aim of the study was to investigate whether the judgment of decisions made by robots differed from the judgment of the same decisions made by a human. Contrary to the expectations, the results showed that decisions of a robot decision maker were not judged differently than decisions of a human expert. This finding contradicts with previous research by Hidalgo et al. (2021) and Malle et al. (2015). There are two possible explanations for this result.

First, the present study did not include the outcomes or consequences of the decision in the scenarios. This is different to the previous studies by Hidalgo et al. (2021) and Malle et al. (2015) in which the consequences were manipulated or explicit. According to Hidalgo et al. (2021), humans were judged by their intentions while robots were judged by the outcomes of their decisions. This means that the judgment of a robot's decision is based on a different aspect than the same decision of a human when the consequences of the decision are known. Since the consequences of the decisions were not known in the present study, this aspect was not available as a basis for judgment. Thus, the judgment of the decision made by the robot decision maker could not be based on the outcomes of the decision.

Second, the scenarios contained a justification of the decision. Traditionally, the decisions of robots included some black-box behavior because they were not able to explain their decisions (Tauchert & Mesbah, 2019). Previous studies already emphasized that the ability to communicate is essential for a positive human-robot interaction (Malle et al., 2015). Furthermore, the results of Tauchert and Mesbah (2019) have shown that advice from robots that are able to communicate may even e preferred over the advice of a human expert. In the present study, the robot was able to explain his decision. Each scenario included a justification of the decision. Therefore, the decision of the robot was no longer a black-box behavior because the robot explained why he chose this decision. In addition, the justifications of the robot and the human decision maker were identical. Thus, the robot had exactly the same possibility of communication as the human.

The present study also showed that the judgment of rational decision differed from the judgment of biased decision depending on the scenario. In the social dilemma, mental accounting, and affect as heuristic scenarios, the rational decision was judged more positively than the biased decision. However, there was no difference between the judgments of the decisions in the sunk cost bias, framing bias, and time discounting scenarios. A possible explanation of the result is that the bias in the scenarios was noticed to different degrees. In the scenarios where no difference in judgment between decisions was found, the bias was not as clear as in the scenarios where a difference was identified. While the bias in the social dilemma scenario, for example, was relatively straightforward to identify as the decision to save the driver, the bias in the framing bias scenario was not so obvious. The decision in the framing bias scenario whether to evacuate half of the people or to save all people with a probability of 50% is a difficult decision to make, in which the judgment of which decision is rational is likely to depend on other factors such as aversion. Therefore, the difference in perceptibility of the bias may have resulted in judgments that were dependent on scenario. In addition to this explanation, no other factor could be identified that distinguished the three scenarios in which a difference between decisions was found from the other scenarios.

Limitations and Future Research

The study had some limitations. First, the study was an online experiment and the participants did not communicate with the decision maker or were actively involved in the situation. Participants might respond differently in a real situation. Other studies showed that the physical presence of a robot has positive impacts on people's trust in robots, their attitudes and also their behavior when compared to a virtual presence of the robot (Huang et al., 2021). Therefore, future studies are needed that study the judgment of decisions made by a robot in a physical setting so that participants can interact and communicate with the robot.

Furthermore, the present study only included the decision but not the outcomes of the decision. Hidalgo et al. (2021) found that robots are judged differently than humans because they are judged by their outcomes while humans are judged by their intentions. Therefore, future studies should analyze whether different outcomes influence the judgment of robot decision makers by including outcomes in their scenarios and manipulate them in different conditions.

Conclusion and Implications

The use of robots is growing and people increasingly rely on the advice of robots. Therefore, it is important to investigate whether people evaluate the decisions of robots in the same way as those of humans. The present research contributed to this field of study by extending the research through using different scenarios. Previous research focused mainly on social or moral dilemmas. In the present study, the scenarios were more widespread. A social dilemma was only one of six scenarios used. Additionally, the robot was given the ability to explain his decision. A justification of the decision was given in the human as well as in the robot decision maker conditions. The results showed that participants did not judge the decisions of a robot differently than those of a human expert. Previous studies have already illustrated that the ability to communicate and to explain a decision can lead to the acceptance of a robot's advice. The results of the present study also provided some indication for the positive effect of the ability to explain a decision on the judgment of a decision made by a robot. Thus, this result had a practical contribution because it showed that under specific conditions, decisions made by robots can be judged in the same way as those of human decision makers. To achieve this, robots should be able to communicate and explain their decisions. In conclusion, participants' judgment did not depend on the decision maker as no difference was found between the robot and the human decision maker, but the judgment of the decision depended on the scenario.

References

Bahner, J. E., Hüper, A.-D., & Manzey, D. (2008). Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *International Journal of Human-Computer Studies*, *66*(9), 688–699. https://doi.org/10.1016/j.ijhcs.2008.06.001

Biller-Andorno, N., Ferrario, A., Joebges, S., Krones, T., Massini, F., Barth, P., Arampatzis, G., & Krauthammer, M. (2021). AI support for ethical decision-making around resuscitation: Proceed with care. *Journal of Medical Ethics*, *48*(3), 175–183. https://doi.org/10.1136/medethics-2020-106786

Bisantz, A. M., & Seong, Y. (2001). Assessment of operator trust in and utilization of automated decision-aids under different framing conditions. *International Journal of Industrial Ergonomics*, *28*(2), 85–97. https://doi.org/10.1016/s0169-8141(01)00015-4

Bless, H., Betsch, T., & Franzen, A. (1998). Framing the framing effect: The impact of context cues on solutions to the ‘Asian disease’ problem. *European Journal of Social Psychology*, *28*(2), 287–291. https://doi.org/10.1002/(SICI)1099-0992(199803/04)28:2<287::AID-EJSP861>3.0.CO;2-U

Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, *352*(6293), 1573–1576. https://doi.org/10.1126/science.aaf2654

Dalal, R. S., & Bonaccio, S. (2010). What types of advice do decision-makers prefer? *Organizational Behavior and Human Decision Processes*, *112*(1), 11–23. https://doi.org/10.1016/j.obhdp.2009.11.007

Dane, E., Rockmann, K. W., & Pratt, M. G. (2012). When should I trust my gut? Linking domain expertise to intuitive decision-making effectiveness. *Organizational Behavior and Human Decision Processes*, *119*(2), 187–194. https://doi.org/10.1016/j.obhdp.2012.07.009

Duxbury, D., Keasey, K., Zhang, H., & Chow, S. L. (2005). Mental accounting and decision making: Evidence under reverse conditions where money is spent for time saved. *Journal of Economic Psychology*, *26*(4), 567–580. https://doi.org/10.1016/j.joep.2004.11.001

Emami, A., Saghafi, F., Zarei, B., Ebrahimzadeh, M., & Davari, A. (2011). Managerial applications of framing effects and mental accounting in network participation. *Advances in Production Engineering & Management*, *6*(1), 57–69.

Emami, A., Welsh, D. H. B., Ramadani, V., & Davari, A. (2019). The impact of judgment and framing on entrepreneurs' decision-making. *Journal of Small Business & Entrepreneurship*, *32*(1), 79–100. https://doi.org/10.1080/08276331.2018.1551461

Evans, J. S. B. T. (2008). Dual-Processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*(1), 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, *40*(2), 351–401. https://doi.org/10.1257/jel.40.2.351

Freedy, A., DeVisser, E., Weltman, G., & Coeyman, N. (2007). Measurement of trust in human-robot collaboration. *2007 International Symposium on Collaborative Technologies and Systems*, 106–114. https://doi.org/10.1109/cts.2007.4621745

Garland, H., & Newport, S. (1991). Effects of absolute and relative sunk costs on the decision to persist with a course of action. *Organizational Behavior and Human Decision Processes*, *48*(1), 55–69. https://doi.org/10.1016/0749-5978(91)90005-e

Gladwell, M. (2007). *Blink: The power of thinking without thinking*. Back Bay Books.

Gonzalez Fabre, R., Camacho Ibáñez, J., & Tejedor Escobar, P. (2020). Moral control and ownership in AI systems. *AI & SOCIETY*, *36*(1), 289–303. https://doi.org/10.1007/s00146-020-01020-z

Goodall, N. J. (2014). Machine ethics and automated vehicles. *Road Vehicle Automation*, 93–102. https://doi.org/10.1007/978-3-319-05990-7_9

Greene, D., Hoffmann, A. L., & Stark, L. (2019). Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning. *Proceedings of the Annual Hawaii International Conference on System Sciences*. https://doi.org/10.24251/hicss.2019.258

Haita-Falah, C. (2017). Sunk-cost fallacy and cognitive ability in individual decision-making. *Journal of Economic Psychology*, *58*, 44–59. https://doi.org/10.1016/j.joep.2016.12.001

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *53*(5), 517–527. https://doi.org/10.1177/0018720811417254

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes*, *70*(2), 117–133. https://doi.org/10.1006/obhd.1997.2697

Hidalgo, C. A., Orghian, D., Albo-Canals, J., de Almeida, F., & Martin, N. (2021). *How Humans Judge Machines*. The MIT Press.

Hodgkinson, G. P., Maule, A. J., Bown, N. J., Pearman, A. D., & Glaister, K. W. (2002). Further reflections on the elimination of framing bias in strategic decision making. *Strategic Management Journal*, *23*(11), 1069–1076. https://doi.org/10.1002/smj.266

Huang, H., Rau, P.-L. P., & Ma, L. (2021). Will you listen to a robot? Effects of robot ability, task complexity, and risk on human decision-making. *Advanced Robotics*, *35*(19), 1156–1166. https://doi.org/10.1080/01691864.2021.1974940

Jenni, K., & Loewenstein, G. (1997). Explaining the identifiable victim effect. *Journal of Risk and Uncertainty*, *14*(3), 235–257. https://doi.org/10.1023/a:1007740225484

Jungermann, H. (1999). Advice giving and taking. *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*. https://doi.org/10.1109/hicss.1999.772750

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, *93*(5), 1449–1475. https://doi.org/10.1257/000282803322655392

Kahneman, D. (2012). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kogut, T., & Ritov, I. (2005). The "Identified victim" effect: An identified group, or just a single individual? *Journal of Behavioral Decision Making*, *18*(3), 157–167. https://doi.org/10.1002/bdm.492

Komiak, S. Y. X., & Benbasat, I. (2006). The effects of personalization and familiarity on trust and adoption of recommendation agents. *MIS Quarterly*, *30*(4), 941–960. https://doi.org/10.2307/25148760

Lieberman, H. (2001). Interfaces that give and take advice. In J. M. Carroll (Ed.), *Human-Computer Interaction for the New Millenium* (1st ed., pp. 475–484). Addison-Wesley Professional.

Loewenstein, G., & Lerner, J. S. (2003). The role of affect in decision making. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 619–642). Oxford University Press.

Malle, B. F., Scheutz, M., Arnold, T., Voiklis, J., & Cusimano, C. (2015). Sacrifice one for the good of many?: People apply different moral norms to human and robot agents. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 117–124. https://doi.org/10.1145/2696454.2696458

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.5465/amr.1995.9508080335

McClure, S. M., Ericson, K. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2007). Time discounting for primary rewards. *Journal of Neuroscience*, *27*(21), 5796–5804. https://doi.org/10.1523/jneurosci.4246-06.2007

Pituch, K. A., & Stevens, J. P. (2016). *Applied multivariate statistics for the social sciences* (6th ed.). Routledge.

Ranyard, R., Hinkley, L., Williamson, J., & McHugh, S. (2006). The role of mental accounting in consumer credit decision processes. *Journal of Economic Psychology*, *27*(4), 571–588. https://doi.org/10.1016/j.joep.2005.11.001

Rossi, F., & Mattei, N. (2019). Building ethically bounded AI. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(1), 9785–9789. https://doi.org/10.1609/aaai.v33i01.33019785

Sanders, T. L., Wixon, T., Schafer, K. E., Chen, J. Y. C., & Hancock, P. A. (2014). The influence of modality and transparency on trust in human-robot interaction. *2014 IEEE International Inter-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*. https://doi.org/10.1109/cogsima.2014.6816556

Schultze, T., Rakotoarisoa, A.-F., & Schulz-Hardt, S. (2015). Effects of distance between initial estimates and advice on advice utilization. *Judgment and Decision Making*, *10*(2), 144–171.

Shinozawa, K., Naya, F., Yamato, J., & Kogure, K. (2005). Differences in effect of robot and screen agent recommendations on human decision-making. *International Journal of Human-Computer Studies*, *62*(2), 267–279. https://doi.org/10.1016/j.ijhcs.2004.11.003

Shneiderman, B. (2007). Human responsibility for autonomous agents. *IEEE Intelligent Systems*, *22*(2), 60–61. https://doi.org/10.1109/mis.2007.32

Shneiderman, B. (2016). The dangers of faulty, biased, or malicious algorithms requires independent oversight. *Proceedings of the National Academy of Sciences*, *113*(48), 13538–13540. https://doi.org/10.1073/pnas.1618211113

Skitka, L. J., Mosier, K., & Burdick, M. D. (2000). Accountability and automation bias. *International Journal of Human-Computer Studies, 52*(4), 701–717. https://doi.org/10.1006/ijhc.1999.0349

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1980). Facts and fears: Understanding perceived risk. *Societal Risk Assessment*, 181–216. https://doi.org/10.1007/978-1-4899-0445-4_9

Starr, N. D., Malle, B., & Williams, T. (2021). I need your advice. . . Human perceptions of robot moral advising behaviors. *ArXiv*. https://doi.org/10.48550/arXiv.2104.06963

Strough, J., de Bruin, W. B., Parker, A. M., Karns, T., Lemaster, P., Pichayayothin, N., Delaney, R., & Stoiko, R. (2016). What were they thinking? Reducing sunk-cost bias in a life-span sample. *Psychology and Aging*, *31*(7), 724–736. https://doi.org/10.1037/pag0000130

Tan, H.-T., & Yates, J. F. (1995). Sunk cost effects: The influences of instruction and future return estimates. *Organizational Behavior and Human Decision Processes*, *63*(3), 311–319. https://doi.org/10.1006/obhd.1995.1082

Tanibe, T., Hashimoto, T., & Karasawa, K. (2017). We perceive a mind in a robot when we help it. *PLOS ONE*, *12*(7), e0180952. https://doi.org/10.1371/journal.pone.0180952

Tauchert, C., & Mesbah, N. (2019). Following the robot? Investigating users' utilization of advice from robo-advisors. *ICIS 2019 Proceedings*. https://aisel.aisnet.org/icis2019/smart_service_science/smart_service_science/8/

Tzioti, S. C., Wierenga, B., & van Osselaer, S. M. J. (2013). The effect of intuitive advice justification on advice taking. *Journal of Behavioral Decision Making*, *27*(1), 66–77. https://doi.org/10.1002/bdm.1790

van den Bosch, K., & Bronkhorst, A. W. (2018). Human-AI cooperation to benefit military decision making. *NATO*. http://resolver.tudelft.nl/uuid:49e7b3ec-602f-4edd-a748-cf88a317e49d

van Swol, L. M. (2011). Forecasting another's enjoyment versus giving the right answer: Trust, shared values, task effects, and confidence in improving the acceptance of advice. *International Journal of Forecasting*, *27*(1), 103–120. https://doi.org/10.1016/j.ijforecast.2010.03.002

Västfjäll, D., Slovic, P., Mayorga, M., & Peters, E. (2014). Compassion fade: Affect and charity are greatest for a single child in need. *PLOS ONE*, *9*(6), e100115. https://doi.org/10.1371/journal.pone.0100115

Volante, W. G., Sosna, J., Kessler, T., Sanders, T., & Hancock, P. A. (2018). Social conformity effects on trust in simulation-based human-robot interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *61*(5), 805–815. https://doi.org/10.1177/0018720818811190

Appendix A

Decision Scenarios Including Decision Options and Justifications of the Decisions

Scenario 1: Sunk cost bias

Scenario

A team has been working on a project for two years already. The project is in the last development stage and it is expected that it will be finished within the next three months. Recent market analyses and customer surveys showed that less customers than expected are interested in the project. Furthermore, it seems that most customers are more interested in the competitor's project. Now it has to be decided whether the project should be continued or terminated.

Decision options

1. The project should be continued. (Biased)
2. The project should be terminated. (Rational)

Justification of the decision

Biased decision

The project team put a lot of time and effort in developing the project for two years. Now, the project is nearly finished. The time as well as the money invested in the project should not get wasted. Therefore, the project should be continued. (1)

Rational decision

The market analyses and customer surveys did not show promising results. Additionally, a competitor is working on a similar project which seems to attract more customer attention. Therefore, the project should be terminated and the remaining budget should be invested in a new, more promising project. (2)

Scenario 2: Framing bias

Scenario

A fire burns in a forest near a small village with 600 inhabitants and threatens to destroy the village. There are two alternatives to try to protect the inhabitants of the village. Option A is to evacuate (part of) the people living in the village. Option B is to fight the fire. The time till the fire reaches the village is very short and only one of these options can be executed. When you evacuate, 300 people will be saved for sure. When you fight the fire, there is a 50% possibility that all inhabitants will be saved.

Decision options

1. Evacuate, which means that 300 people will be saved. (Biased)
2. Fight the fire, which means that there is a probability of 50% that all inhabitants will be saved. (Rational)

Justification of the decision

Biased decision

When you evacuate it is certain that 300 inhabitants will be saved whereas when fighting the fire there is only a 50% possibility that all inhabitants will be saved. Since Option A offers a certain survival chance, this option should be chosen and 300 people should be evacuated. (1)

Rational decision

By evacuating, it is certain that half of the inhabitants will be saved. Option B on the other hand offers a 50% possibility that all inhabitants will be saved. Since Option B offers the chance to save all people, this option should be chosen and the fire should be fought. (2)

Scenario 3: Social dilemma

Scenario

Imagine that one person drives in an autonomous car through a city. Suddenly three pedestrians cross the street. There is no chance for the autonomous car to avoid colliding with the pedestrians without crashing into a tree on the side of the road and thus, risking the life of the driver. In order to save the life of the driver, the autonomous car would have to run over the pedestrians. For designing purposes it has to be decided what the autonomous car should do in this situation.

Decision options

1. The autonomous car should avoid the collision with the three pedestrians and risks the life of the driver. (Rational)

2. The autonomous car should save the life of the driver and run over the three pedestrians. (Biased)

Justification of the decision

Biased decision

Autonomous cars should protect the life of the driver at all cost. Even if the lives of the pedestrians have to be risked, the autonomous car should save the life of the driver. Therefore, the autonomous car should not avoid the collision but should run over the pedestrians instead. (2)

Rational decision

Since there is only one person in the car but three pedestrians on the street, it would be better to save the lives of the three pedestrians. The option that saves most lives should be chosen. Therefore, the decision is that the autonomous car avoids the collision with the pedestrians even if the life of the driver is risked by choosing this option. (1)

Scenario 4: Mental accounting

Scenario

Two tickets have to be bought for an event. The salesman in the ticket shop tells that the price for the tickets is 100€. Moreover, he shares the information that they have other tickets for the same event in another of their ticket shops. These tickets would cost only 95€ but the walk to the other shop would take 10 minutes. The salesman wants to know which tickets will be bought.

Decision options

1. Buy the tickets for 100€. (Biased)
2. Walk 10 minutes to the other shop and buy the tickets for 95€. (Rational)

Justification of the decision

Biased decision

In the other ticket shop, 5€ could be saved. However, a walk of 10 minutes is necessary and 5 euros is only 5% of the total prices. This relatively small amount is not worth the effort so the tickets for 100€ should be bought. (1)

Rational decision

In the first ticket shop, the tickets are 5€ more expensive than in the other ticket shop. To save the 5€ it is only necessary to walk 10 minutes to the other ticket shop. Saving 5 euros this way is the most rational thing to do, so the tickets for 95€ should be bought. (1)

Scenario 5: Affect as heuristic

Scenario

You are able to donate 100€ to one of two nonprofit organizations. The first organization *Save Africa* distributes food and water in regions of Africa that are affected by drought. The organization helps many people that are suffering from the conditions. The second organization *helpCHILDREN* rescues children in crisis regions. One of these children is Chad. He is a 6 years-old boy and his village was destroyed by a hurricane. With your donation, the organization can provide Chad with food and water for several weeks. The organization for the donation has to be chosen now.

Decision options

1. Donate the money to *Save Africa* and aid many people in regions affected by drought. (Rational)
2. Donate the money to *helpCHILDREN* and provide help to Chad. (Biased)

Justification of the decision

Biased decision

If the donation goes to *helpCHILDREN* they will be able to care for Chad for several weeks. Knowing that he is still a little boy and needs the donation to survive, *helpCHILDREN* should be chosen. (2)

Rational decision

If the donation goes to *Save Africa*, they will be able to help different families and people that are affected by the drought in the regions and are starving. Since more people could be helped if the donation goes to *Save Africa*, not only the children, this organization should be chosen. (1)

Scenario 6: Time discounting

Scenario

Your superior is thinking about paying you a bonus and offered you two options. Option A is that you would get a bonus of 300€ immediately. Option B is that you would get a bonus of 350€ at the end of the year which is a delay of 6 months. Now your superior needs your decision.

Decision options

1. Go for the immediate bonus of 300€. (Biased)
2. Wait half a year and receive a bonus of 350€. (Rational)

Justification of the decision

Biased decision

Although the delayed bonus is higher than the immediate bonus, the immediate bonus is the safer option. It is better to be certain to have the bonus now than the delayed bonus after half a year. Therefore, the immediate bonus should be chosen. (1)

Rational decision

The delayed bonus is higher than the immediate bonus. Even if the inflation rate is subtracted, the delayed bonus at the end of the year is still higher. Therefore, the delayed bonus should be chosen. (2)

Appendix B

Items Measuring the Demographics, Dependent and Control Variables

Demographic Questions

What is your age? (free text box)

Which gender do you have? (male, female, non-binary / third gender)

What is your nationality? (free text box)

Dependent Variables (After Each Scenario)

All items measured on a 5-point Likert scale (1 = strongly disagree, 2 = disagree, 3 = neither agree or disagree, 4 = agree, 5 = strongly agree)

Instructions

Please indicate to what extent you agree to the following statements. Select the answer that applies to you best in this moment.

Items

1. A correct decision is made.
2. The decision is morally acceptable.
3. I trust the decision maker.
4. I like the decision maker.
5. I would have done the same if I were in this situation.
6. The decision is sensible.
7. The other decision option would have been the better alternative.*
8. The decision maker should be replaced by a [robot / human]. *[different]**
9. The decision maker should be replaced by another [robot / human]. *[same]**

* Item has been recoded

Control Variable

Item 1 measured on a 5-point Likert scale (1 = nothing at all, 2 = almost nothing, 3 = sometimes, 4 = often, 5 = a lot)

Items 2 to 9 measured on a 5-point Likert scale (1 = strongly disagree, 2 = disagree, 3 = neither agree or disagree, 4 = agree, 5 = strongly agree)

Instructions

The survey consists of 9 statements. Please read every statement carefully. There are no right or wrong answers, just answer honestly. Select the answer that applies to you best in this moment.

Items

1. I heard about artificial intelligence (AI).
2. I am worried when I think of AI.*
3. In my opinion the AI has more benefits than risks.
4. I perceive that AI creates new problems.*
5. The risks of AI are higher than the benefits.*
6. When I think of AI, I am hopeful.
7. I am angry when I think of AI.*
8. In my opinion AI solves problems.
9. I think that the risks of AI equal the benefits.

* Item has been recoded

Appendix C
Informed Consent and Debriefing

Informed consent

Dear participant,

Thank you for your participation. The purpose of the present study is to determine how people evaluate various decisions in different scenarios. We collect the data as part of my master's thesis at the University of Twente. The study should not take longer than 20 minutes to complete. The study starts with a survey including some demographic questions (age, gender, nationality). The survey is followed by different scenarios after each you will be asked to rate the decision(maker) by answering some questions about the scenario. Additional information about the scenario and decisions will be provided. After the last scenario, you will be presented with a survey including additional questions related to the study.

Please read the following information carefully.

Your anonymity will be ensured at any point in the study. All data will be processed confidentially and anonymously and cannot be disclosed in personally identifiable way. You can participate in the study if you are at least 18 years old. Furthermore, your participation is voluntary. You can withdraw from the study at any time without any reason. There are no foreseeable risks from participating in the study and the study is in line with all ethical norms. Your data will not be accessed by any third party. Your data will only be used for academic purposes in the present study if you agree to the informed consent.

For additional information and questions, please contact me via e-mail (e-mail).

If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher, please contact the Secretary of the Ethics Committee of the Faculty of Behavioral, Management and Social Sciences at the University of Twente (ethicscommittee-bms@utwente.nl).

If you agree to participate in the study, you confirm that you read the information and understand that your anonymized data will be used for academic purposes.

- I agree to participate in the study
- I disagree to participate in the study

Debriefing

Dear participant,

Thank you that you have completed the study. Before the study, we informed you that the purpose of the study is to determine how people judge the advisor and the decision made in different situations. The study also has a second purpose that we could not tell you before the study. The second purpose is to investigate whether there is a difference in the judgment if the decision maker is a robot or a human. Therefore, in the scenarios either a human or a robot decision maker was described. However, the robot and the human advisor made the same decision and justifications were presented in the conditions.

For additional information or questions about the study, if you want to withdraw from the study after you were informed about the actual purpose of the study or if you are interested in the results of the study, please contact me via e-mail (e-mail).

If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher, please contact the Secretary of the Ethics Committee of the Faculty of Behavioral, Management and Social Sciences at the University of Twente (ethicscommittee-bms@utwente.nl).

Appendix D

Introduction

Introduction in the robot decision maker condition

The study consists of six different decision tasks. Each task will start with a short scenario. At the end of the scenario, you will be presented with two decision options. Additionally, a robot has been consulted to make the decision. The robot is an artificial intelligence (AI) system and includes an algorithm that analyzes all information of a situation. Additionally, the robot analyzes the decision options and evaluates information and outcomes of past situations that were similar to the present scenario. Based on all information, the AI system evaluates which decision option would be the best to achieve a preferrable outcome. The robot is regularly used to support humans by making decisions. You will be presented with the decision of the robot as well as his justification of the decision. Afterwards you will be asked to indicate your opinion about the robot's decision by answering some questions related to the decision made in the scenario.

Before the first decision task, you will be asked to answer some demographic questions. Additionally, after the last decision task, you will be asked to answer some additional questions related to the study.

Introduction in the human decision maker condition

The study consists of six different decision tasks. Each task will start with a short scenario. At the end of the scenario, you will be presented with two decision options. Additionally, another person has been consulted to make the decision. The person is an expert in different fields and is often consulted to make decisions in critical situations. Therefore, the person has a lot of experience in making decisions. Based on the experience from other situations and the knowledge about the outcomes in these situations, the expert tries to choose the options that would be the best to achieve a preferrable outcome. You will be presented with the decision of the person as well as his justification of the decision. Afterwards you will be asked to indicate your opinion about the person's decision by answering some questions related to the decision made in the scenario.

Before the first decision task, you will be asked to answer some demographic questions. Additionally, after the last decision task, you will be asked to answer some additional questions related to the study.

Appendix E

Rotated Factor Matrix

Table E1

Rotated Factor Matrix of the Dependent Variables in Scenario 1

Item	Factor loading	
	1	2
1. A correct decision is made.	.89	
2. The decision is morally acceptable.	.71	
3. I trust the decision maker.	.76	
4. I like the decision maker.	.68	
5. I would have done the same if I were in this situation.	.89	
6. The decision is sensible.	.70	
7. The other decision option would have been the better alternative.*	.71	
8. The decision maker should be replaced by a human / robot.*		-.78
9. The decision maker should be replaced by another robot / human.*		.66

Note. Only factor loadings higher than .30 are displayed. Varimax Rotation with Kaiser Normalization.

* Item has been recoded

Appendix F

Correlation Matrix

Table F1

Pearson correlations for the demographics, attitude towards AI and the means of the dependent variables

	1	2	3	4	5	6
1 Age	-					
2 Gender	-.08	-				
3 Nationality	-.21	.03	-			
4 Attitude towards AI	-.25*	-.12	.24*	-		
5 Mean Judgment	-.20	-.02	-.20	-.09	-	
6 Mean (Perceived) Appropriateness	.14	.08	.08	-.02	.26*	-

* The correlation is significant at the level $p < .05$.