Szeli, Leon

Article

UX in AI: Trust in Algorithm-based Investment Decisions

Junior Management Science (JUMS)

Provided in Cooperation with:
Junior Management Science e. V.

Suggested Citation: Szeli, Leon (2020) : UX in AI: Trust in Algorithm-based Investment Decisions, Junior Management Science (JUMS), ISSN 2942-1861, Junior Management Science e. V., Planegg, Vol. 5, Iss. 1, pp. 1-18, https://doi.org/10.5282/jums/v5i1pp1-18

This Version is available at:
https://hdl.handle.net/10419/294920

Junior Management Science 5(1) (2020) 1-18

Junior Management Science

journal homepage: www.jums.academy

UX in AI: Trust in Algorithm-based Investment Decisions

Leon Szeli

University of Cambridge and Technical University Munich

Abstract

This Thesis looks at investors' loss tolerance with portfolios managed by a human advisor compared to an algorithm with different degrees of humanization. The main goal is to explore differences between these groups (Humanized Algorithm, Dehumanized Algorithm, Humanized Human and Dehumanized Humans) and a potential diverging effect of humanizing. The Thesis is based on prior research (Hodge et al., 2018) but incorporates new aspects such as additional variables (demographics, prior experiences) and a comparison between users and non-users of automated-investment products. The core of this research is an experiment simulating an investment portfolio over time with four different portfolio managers. Subjects were asked to decide if they want to hold or sell a declining portfolio at five points in time to measure their loss tolerance. A cox regression model shows that portfolios managed by the Humanized Human had the highest loss tolerance. Humanizing leads to higher loss tolerance for the human advisor but to lower loss tolerance for algorithmic advisors within the non-user group.

Keywords: Künstliche Intelligenz; Artificial Intelligence; Behavioral Finance; Behavioral Economics; Human-Computer-Interaction; User Experience; Investmententscheidungen; Nutzervertrauen.

1. Introduction

This world is endowed with a limited number of AI (Artificial Intelligence) experts. The average citizen has a cartoonish understanding of AI based on overindulgent media headlines, which in itself is a by-product of a rapidly developing area. These headlines can have positive or negative connotations. Headlines such as the following are abundant in some of the most prestigious newspapers: "News broadcast triggers Amazon Alexa devices to purchase $170 doll houses", "Robot passport checker rejects Asian applications because eyes are closed", "Robots judge a beauty contest and don't select women with dark skin", or "Microsoft's Twitter chatbot turns anti-feminist and pro-Hitler" (Leaden, 2017). Not all of these examples are caused by AI or algorithmic flaws, but in the general perception that does not matter. These headlines fuel mistrust and are not easily remedied by positive reports.

Nonetheless, trust in AI will be the driving factor for adoption in many industries. A good example to illustrate this is the financial sector. While AI already enjoys a high level of trust in some areas (entertainment, navigation), only half of people trust algorithmic investment advice (Shandwick, 2016). If the trust level does not change, Roboadvisors will never be able to reach mass market. This fact in itself makes for a compelling research project. Whereas, for the past ten years, most researchers have agreed that humans trust humans more than AI. A recent range of experiments has contradicted that common understanding (Hodge et al., 2018).

Previously, you received advice from a human being, but now, due to rapid technological progress, you can also attain advice from AI systems like Roboadvisors. Roboadvisors are automated, algorithmic investment products. In 2017, Roboadvisors already had $9.1 billion under management globally (Eule, 2017). Experts predict $2.2 trillion under Roboadvisor management by 2020 (Epperson et al., 2015). Whether or not these predictions become true depends on whether users adopt the new technology. Whether they take the financial advice of a human being over an algorithm critically depends on trust (Hodge et al., 2018).

Yet, while the Roboadvisor industry is growing, there is little research examining how and when consumers trust the financial advice of algorithms. Which factors, from a psychological point of view, influence whether consumers trust AI? Despite the vast number of factors influencing trust and the variety of AI use cases, this paper focuses only on a selection of them. One factor that could drive trust in these products is the humanization of the algorithm. Does it help to make the algorithm more human and social, e.g. by naming it? AI-based assistants such as Amazon's Alexa, Microsoft's Cortana and Apple's Siri have a humanized interface, although

DOI: https://dx.doi.org/10.5282/jums/v5i1pp1-18

contemporary research voices doubt the positive influence on trust (Hodge et al., 2018). From this paradox, the research question of this thesis is derived: Does trust (measured as loss tolerance) in the financial advice of an algorithm (versus a human) depend on the degree to which the algorithm is humanized (giving the algorithm a human name)? To answer the research question, we conducted an online survey including experiments with users of the Roboadvisor Ginmon and non-users. Moreover, we looked at additional factors such as demographics and prior investment experiences.

2. State of Research

This chapter sheds light on the current state of research regarding trust. First, we look at the trust in humans – its original sense. Subsequently, trust in technology is examined.

2.1. Defining Trust

Trust as a basis for decision-making in diverse contexts has been studied in various fields. Cho et al. (2015) provide a clear overview of the multidisciplinary meanings of trust (Cho et al., 2011; Gambetta, 2000; James, 2002; Kydd, 2005; Lagerspetz, 1998; Lee and See, 2004; Mayer et al., 1995; Rotter, 1980):

Based on the multitude of meanings of trust, Cho et al. (2015) summarize the concept of trust as follows:

> "Trust is the willingness of the trustor (evaluator) to take risk based on a subjective belief that a trustee (evaluatee) will exhibit reliable behavior to maximize the trustor's interest under uncertainty (e.g., ambiguity due to conflicting evidence and/or ignorance caused by complete lack of evidence) of a given situation based on the cognitive assessment of past experience with the trustee." (p. 28:5)

When individuals decide if they trust an entity, we call this process trust assessment. Cho et al. (2015) summarize this process as "Trustor i assesses Trustee j's trust if j can perform Task A" (p. 28:5). According to game theoretic approaches, trustor i is defined as someone who maximizes their interest (or utility) from the relationship with the trustee (Chin, 2009). Trustee j is someone who can cause impact on a trustor's utility with his/her behavior (Castelfranchi and Falcone, 2010).

Task A is a crucial factor in the decision if i trusts j as the importance of the tasks influences the risk assessment and potential outcomes. Afterwards, i adjusts trust according to whether the decision was right or wrong (Cho et al., 2015).

2.2. Trust in Information Technology

This chapter explains trust between humans and information technology. The concept of trust is not only applicable to human-to-human relationships but also to human-to-technology relationships (Coeckelbergh, 2012). Nevertheless, there are differences between the two relationships (Mcknight et al., 2011). Firstly, regardless whether it is about trust in people or trust in technology, it involves risk and uncertainty. In the case of humans, you lack total control. You depend on the trustee to fulfill expected responsibilities which s/he might not fulfil – (un)intentionally. In the case of machines, you also do not have the control as a user since the technology might not demonstrate the expected capability (i.e., without intention). For example, when you trust Dropbox to save your data, you are exposed to the risk of data transmission over the internet and storing confidential data on a third-party server. Secondly, people have moral agency and volition whereas technology is amoral and non-volitional. For example, when you use a word processing program like Grammarly it will correct misspelled words and grammatical errors. But a benevolent human copyeditor might alter your text (reflecting his/her willingness and therefore volition) in order to help you improve even though it is not part of his/her job. A technology will only follow instructions programmed. Thirdly, the trustor's expectations regarding the object of dependence might be different when comparing humans to machines. When trusting people, you expect them to fulfil a task for you in a competent way. When trusting a technology, you want it to demonstrate possession of functionality. A human helps you if s/he cares for you and is benevolent towards you. A machine's helpfulness is not rooted in moral agency, but you still expect effective help (e.g. a help menu). When trusting humans, we hope for integrity, reliability and consistence in their actions. Due to humans' free will, this is a risk. In a machine, we are looking for reliability. It should operate consistently without failing. Potential failures are caused by a bug and not by deliberate actions (Mcknight et al., 2011).

As described in chapter 2.2.3, trust in technology can be influenced by many factors. Siau and Wang (2018) structures these in multiple dimensions: Human characteristics (Personality, Ability), Environment Characteristics (Culture, Task, Institutional Factors) and Technology Characteristics (Performance, Process, Purpose) (Siau and Wang, 2018). For example, someone with a rather trusting personality (Human Characteristic) is likely to trust the technology in the task of filesharing (Environment Characteristic) no matter whether Google Drive or Dropbox is used (Technology Characteristic).

2.3. Trust in Automated and Digital Domains (e-Trust)

So far, we have been describing trust as a general concept and in the context of Information Technology. This chapter focuses on trust in more digital and automated domains.

2.3.1. Defining Trust in AI, HCI and Automation

Cho et al. (2015) defined key trust components for the three domains AI (Artificial Intelligence), HCI (Human-Computer Interaction) and Automation. All of the three applications fall in the category of "e-trust" which is trust occurring in digital contexts (Taddeo, 2010). They are structured based on four kinds of trust. First, Communication Trust which can be measured objectively (i.e. network connectivity). Second, Information Trust which includes quality

Table 1: Multidisciplinary Definitions of Trust according to Cho et al. (2015)

Discipline	Meaning of Trust	Source
Sociology	Subjective probability that another party will perform an action that will not hurt my interest under uncertainty and ignorance	Gambetta (2000)
Philosophy	Risky action deriving from personal, moral relationship between two entities	Lagerspetz (1998)
Economics	Expectation upon a risky action under uncertainty and ignorance based on the calculated incentives for the action	James (2002)
Psychology	Cognitive learning process obtained from social experiences based on the consequences of trusting behaviors	Rotter (1980)
Organizational Management	Willingness to take risk and being vulnerable to the relationship based on ability, integrity, and benevolence	Mayer et al. (1995)
International Relations	Belief that the other party is trustworthy with the willingness to reciprocate cooperation	Kydd (2005)
Automation	Attitude that one agent will achieve another agent's goal in a situation where imperfect knowledge is given with uncertainty and vulnerability	Lee and See (2004)
Computing & Networking	Estimated subjective probability that an entity exhibits reliable behavior for particular operation(s) under a situation with potential risks	Cho et al. (2011)

of information and credibility. Third, Social Trust which describes trust between humans in a social network. And fourth, Cognitive Trust which refers to accumulated knowledge regarding reliability and competence of the trusted party.

As you can see from Table 2, there are many overlaps in the different areas of trust and applications (availability in Communication Trust, belief in Information Trust, importance in Social Trust, expectation in Cognitive Trust). Nonetheless, there are differences and you have to be sure if you are talking about just one of the domains, an overlap of two or a mixture of multiple (Cho et al., 2015).

2.3.2. Human vs. Machine: Algorithm Aversion vs. Algorithm Appreciation

As initially stated, algorithms perform better than humans in many cases. Nonetheless, many humans rather trust a human prediction than an algorithmic prediction (Diab et al., 2011; Eastwood et al., 2012). In literature, this phenomenon is called algorithm aversion. Experiments support that – even if the algorithm outperforms the human (Dietvorst et al., 2015). Humans are more tolerant if a human is mistaken than if it is an algorithm (Dietvorst et al., 2018). Additionally, humans put more weight on human statements (Önkal et al., 2009; Promberger and Baron, 2006).

Multiple scholars investigated potential reasons for algorithm aversion: The human desire for perfection (Dawes, 1979; Einhorn and Hogarth, 1988; Highhouse, 2008), the assumption that the human learns based on past experience (Highhouse, 2008), the missing humaneness (Dawes, 1979; Grove and Meehl, 1996) and ethical concerns (Dawes, 1979).

Algorithm aversion has been the status quo, but a recent set of experiments conducted by researchers from Harvard Business School contradicts these findings (Logg et al., 2019). They found empirical evidence for Algorithm Appreciation which means that the humans relied more on algorithmic advice than on human advice. They ran six experiments with the following findings: a) people relied on algorithmic advice rather than on humans when estimating a person's body weight based on a picture (Experiment 1A), forecasting the popularity of songs (chart ranking) and romantic matches (attractiveness of the opposite gender) (Experiments 1B and 1C), b) algorithm appreciation persisted no matter if the advice appeared jointly or separately (Experiment 2) and c) Algorithm appreciation was reduced when people could choose between an algorithm's estimate and their own (versus an external advisor's; Experiment 3) and when they had experience in forecasting (Experiment 4).

As a conclusion, we can say that there is no clear answer to the questions whether humans prefer to trust humans or algorithms. There is additional research required in all kinds of domains.

2.3.3. Relevant Trust Factors in AI-based Products

Various factors influencing trust have been identified in past research. The last decades were mainly focused on trust in automation and robotics. In the past years, also studies about trust in digital products and, more specifically, AI have been published. To get an overview, the identified factors are summarized in this chapter. Tables with all factors can be found in the appendix (Table 13, 14 and 15). The sources are mainly meta analyses which represent multiple other papers (Adams et al., 2003; Cho et al., 2015; Hancock et al., 2011; Siau and Wang, 2018). Factors which are not applicable for Roboadvisors were neglected. Highly overlapping factors and factors which are only differing due to denota-

Table 2: Key trust components in different domains according to Cho et al. (2015)

	Communication Trust	Information Trust	Social Trust	Cognitive Trust
Artificial Intelligence	Cooperation, availability	Belief, experience, uncertainty	Importance, honesty	Expectation, confidence, hope, fear, frustration, disappointment, relief, regret
Human-Computer Interaction	Reliability, availability	Belief, experience, uncertainty	Importance, integrity	Expectation, frustration, regret, experience, hope, fear
Automation	Reliability, availability	Belief, experience	Integrity, importance	Expectation, confidence

tion of the authors were summarized as one. The factors can be separated into three categories: Factors regarding the Trustee (the institution/technology (provider) – in our case Ginmon/Ginmon's algorithm), Trustor (the human user) and Environment (contextual, situational) (Adams et al., 2003). To summarize, we can say that the Trustor is looking for a reliable and adaptable product that is transparent in its actions. The Trustee's trust in AI depends on his/her demographics, personality and prior touchpoints with the topic. The environmental factors concern the task, legal aspects, risk as well as influence from other people or the current state of mind. Due to the multitude of trust factors, we decided to focus on one which seems particularly interesting: humanization. This decision, as well as the factor itself, will be elaborated further in chapter 2.4.4 and 2.4.5.

Researchers came up with a wide range of different theoretical models regarding trust in technology and/or automation incorporating some of the mentioned factors. The main frameworks which were developed between 1989 and today are summarized in a Table 12 in the appendix.

We can observe the following patterns among trust models: a) most focus on automation and include an operator that receives recommendations but still has the decision power, b) numerous factors are involved due to the complexity of trust as a concept and c) most models are generic and not tailored towards a specific technology (e.g. algorithmic investment decisions).

There is no model (yet) whose application would help us answer our research question on AI-based investment products. Therefore, we follow a rather exploratory approach which is not based on a theoretical model but nonetheless incorporates empirical findings from the past.

2.4. Trust in Automated Investment Decisions

Trust plays an important role in finance, no matter if it is on the global, institutional level or on the people-to-people level (Bottazzi et al., 2016). Giving another entity money involves risk and vulnerability. For digital finance products, trust is even more crucial since there is less human face-to-face interaction (Greiner and Wang, 2010). Especially, new and innovative products (fintechs) are key as there is less experience value (Van Thiel and Van Raaij, 2017).

2.4.1. Trust in Roboadvisors

This research is focusing on so-called Roboadvisors which came up after the financial crisis in 2008. "Robo-Advisors provide investing advice, wealth management services, sometimes in addition to data aggregation. These Fintech companies provide investment advice and trading services that are automated using algorithms and artificial intelligence" (Gold and Kursh, 2017, p.140). In 2017, Roboadvisors managed $200 billion in assets (Eule, 2017). Roboadvisors have two main benefits: cost saving due to automation and transparency due to the user interface (Salo, 2017). Therefore, also people who did not have the financial resources to pay a human financial advisor to manage their money have the chance to invest in a comparable manner. There is no human intervention other than the user deciding to liquidate the portfolio managed by the algorithm. How can trust be built if there is no human involved? A similar case is e-banking where ease of use, usefulness, perceived privacy and perceived security are considered trust builders (Yousafzai et al., 2003). In the case of Roboadvisors, price and trustworthiness (initial and ongoing trust) additionally come into play when it comes to building trust (Lee et al., 2018).

Roboavisors have different customer segments with different expectations. Salo (2017) identified four groups of Roboadvisor users based on their technoliteracy and financial literacy. The "delegators" want to outsource their investing to someone. Therefore, they are looking for an easy, available, neutral solution with low costs. The "optimizers" look for the most efficient (easier/cheaper) solution. They usually have a little more knowledge and net worth than the delegators and would like to be involved in the investment process, if it could increase returns. The third group is called "Do-It-Yourself". These users want to make investment decisions themselves and are only seeking for advice to consider. The advisor plays a smaller role than in the other two groups (Carré et al., 2016; Salo, 2017). When we talk about building users trust, we have to be clear if we are talking about

all potential users, current users or a specific user group. Since there is only very little research on users of Roboadvisors present, this work is looking at all potential users as well as current users. Looking at a specific user group might be a good idea in a second step.

2.4.2. Trust as a Reaction to Financial Losses

As described in earlier chapters, trust is closely related to risk: "Trust is the willingness of the trustor (evaluator) to take risk [...]." (Cho et al., 2015, p 28:5). Nonetheless, it is not the same (Houser et al., 2010). Traditional finance theory and psychological literature have different understandings of risk. In traditional finance theory, risk is defined as the variance of the expected distribution of returns (Bernstein, 1996; Haugen, 1995). For psychologists, risk is synonymous to loss. The higher your trust in the decision-maker, the bigger your loss tolerance (Payne, 1975; Shapira, 1995; Slovic and Lichtenstein, 1968; Teigen, 1996). Also, you are more likely to forgive a mistake (which is a bad trade in the case of Roboadvisors) (Dietvorst et al., 2018). In this thesis, we take the psychological perspective.

In addition to risk, vulnerability was mentioned multiple times in the different definitions of trust. Trust is only required, if you have "something to lose". Furthermore, trust is an important factor in investment decisions, such as buying and selling stocks (Guiso et al., 2008). Therefore, we decided to simulate an investment portfolio in our experiment, which will be described later. To find out how the respondents react to losses, we decided to have a constantly declining portfolio. We should keep in mind that we are not measuring trust directly but loss tolerance (Will they liquidate the portfolio?)[1].

Behavioral Economics literature suggests that two biases might occur in this experiment: loss aversion and sunk costs. According to prospect theory (Kahneman and Tversky', 1979), individuals prefer avoiding losses to making equal gains. In the context of stock markets, this can lead to holding stocks which are losing in value for too long instead of selling them despite loss (Odean, 1998). Another effect that leads to following through with your portfolio, even if it is declining, is called sunk costs. Sunk costs are costs that have already been incurred and cannot be reversed. In our case it means that investors already have spent time (e.g. keeping track of portfolio's value). Additionally, costs that will irrevocably incur in the future are included (Arkes and Blumer, 1985). Since the portfolio of our participants in our experiment is declining and involves sunk costs, we have to keep the two biases in mind.

2.4.3. The Case of Ginmon

This work looks at one Roboadvisor – Ginmon – in specific. The Frankfurt-based startup was founded in 2014 and is among the leading Roboadvisors in Germany. They offer an automated investment portfolio that is managed by an algorithm called Apeiron. Ginmon charges you 0,39% of your invested amount and 10% of your gains.

Your portfolio is allocated based on your risk class (from 1 to 10). To determine your risk class, you have to answer 9 questions – just like with a human advisor. The questions can be found in the attachment. After being assigned to a risk class there is no other human intervention needed. The only intervention possible is liquidating the portfolio. The algorithm takes care of everything else (buying and selling, rebalancing). The customer can check the current value of the portfolio at any time.

The minimum investment amount is 1000€ , which is much less than traditional advisors require. Nonetheless, wealthy individuals are more attractive customers as they invest more money and therefore the share Ginmon receives is bigger. Generally speaking, people with high net worth are older (Krause and Schäfer, 2005). Also, these people have more reservations towards digital products (Millward, 2003). This paradox explains the company's interest in the question how to build trust in their product.

2.4.4. Anthropomorphism

As mentioned earlier, humanization of algorithms or depicting them as social can influence users' trust. The technical term for that phenomenon is called Anthropomorphism. According to Guthrie (1993) Anthropomorphism describes perceiving humanlike characteristics such as physical appearance, emotional or mental states and motivation in nonhuman agents (Guthrie, 1993). Humanizing nonhuman agents in order to increase user acceptance has been investigated in the past (André et al., 2018; Epley et al., 2008). Examples include making a chatbot more social by imitating human-like behavior, giving it a face/avatar or a name.

2.4.5. Research Question & Hypotheses

Humanizing technology demonstrated benefits in some domains (e.g. IBM's "Watson" or Amazon's "Alexa") but the effects are not clear in financial advising (Hodge et al., 2018). Human-Computer-Interaction research in the past suggests that humanization of technology leads to positive feelings towards the technology and increases the likelihood to use the technology (Burgoon et al., 2000; Chaminade et al., 2007; Eyssel et al.; Gong, 2008; Venkatesh, 2000). However, that does not necessarily mean that the technology is perceived more trustworthy or persuasive (Nan et al., 2006; Riegelsberger et al., 2005). It has been shown that naming a technology can decrease its credibility, which also decreases trust. A technology can be perceived simple and unable to complete complex tasks if it is named (Hafer et al., 1996; Riegelsberger et al., 2005). Naming a human on the other hand, is beneficial. Sharing personal information increases the trust level because the advisor is willing to risk his reputation (Garner, 2005; O'Keefe, 1990; Pornpitakpan, 2004).

Hodge et al. (2018) have shown that humanizing a Roboadvisor decreases the likelihood of subjects to follow

[1] Limitations caused by the dependent variable are addressed in Chapter 5. Interpretation, Limitations and Future Research

investment recommendations. For human advisors the opposite is the case. The Humanized Human was the advisor whose recommendations were followed the most (73,6%). Second was the Dehumanized Roboadvisor (69,7%), third the Humanized Roboadvisor (59,2%) and fourth the Dehumanized Human (54,7%) (Hodge et al., 2018).

We are interested in the threshold where people lose trust (which means sell) in their portfolio – their loss tolerance. Is there a difference in users' trust depending on the type of advisor (human/algorithm) and the level of humanization (humanized/dehumanized)?

There are many definitions of losing trust but in our case, we understand it as a point in time where the trustee (user) loses hope in the ability of the trustor (Roboadvisor) to reach the desired outcome (increase in portfolio value)[2]. We assume that each individual has a certain "pain threshold" of loss and as soon that is exceeded, they lose trust and sell (liquidate the portfolio). Based on the mentioned findings, we formulate our hypotheses:

H1: Respondents' loss tolerance is the highest for portfolios managed by Humanized Humans ("Anlageberater Charles").
H2: Respondents' loss tolerance is the second highest for portfolios managed by Dehumanized Algorithms ("Algorithmus").
H3: Respondents' loss tolerance is the third highest for portfolios managed by Humanized Algorithms ("Algorithmus Charles").
H4: Respondents' loss tolerance is the lowest for portfolios managed by Dehumanized Humans ("Anlageberater").
H5: Respondents' loss tolerance is higher for portfolios managed by Humanized Human than for a Dehumanized Human.
H6: Respondents' loss tolerance is higher for portfolios managed by Dehumanized Algorithm than for a Humanized Algorithm.

While our own hypotheses are mainly based on Hodge's findings, there are some distinct differentiations to his work. Firstly, Hodge focuses on investors' recommendations and whether people follow advice. For Roboadvisors, you usually just invest money and the Roboadvisor has full control afterwards. This means that the user is not involved in decision-making anymore due to automation and therefore does not receive advice. As a result, we decided to focus on the question if and how long people trust the algorithm with their money instead of "following recommendations". Secondly, Hodge's sample consists of 108 MBA students which means they have a) a lot of background knowledge in finance and b) are rather young. Reaching older users with less knowledge about digital products and finance is crucial for mass adoption of Roboadvisors. Therefore, we were aiming for a more diverse sample. Thirdly, Hodge gave participants extensive background information about the potential investment decisions. AI-based products and Roboadvisors specifically cannot provide this in real life and, therefore, "blind" trust is required from users. As a consequence, we decided to give subjects less background information. Fourth, Hodge is looking at one-time investment decisions while we are looking at a portfolio over time. Fifth, Hodge et al. (2018) analyse their data with t-tests and ANOVAs while we took a different approach for the data analysis (cox regression). Sixth, Hodge et al. (2018) did not look at demographic factors or prior investment experiences which we do by controlling for them and comparing users of automated investment products to non-users. We try to thereby contribute to validating the findings while also expanding them.

3. Method

This chapter provides additional information on the sample and survey participants, data collection procedure, and the instruments.

3.1. Sample and Participants

The survey, conducted in December 2019, was sent to German Ginmon customers as well as non-customers. The goal was to gain insights on users and non-users of AI-based products. A total of 258 people took part in the survey with an almost equal split between users and non-users. 82% are male, 18% female. The average age was 36,6, the youngest participant was 18, the oldest 80. When interpreting results, we have to keep in mind that this sample does not represent the German population[3].

3.2. Data Collection Procedure

The survey contained 13 closed questions – and took about five minutes to complete. Ginmon users were contacted using Ginmon's newsletters. The non-users were mainly recruited using the authors' social media profiles. These efforts lead to 258 started surveys out of which 223 were completed.

3.3. Instruments

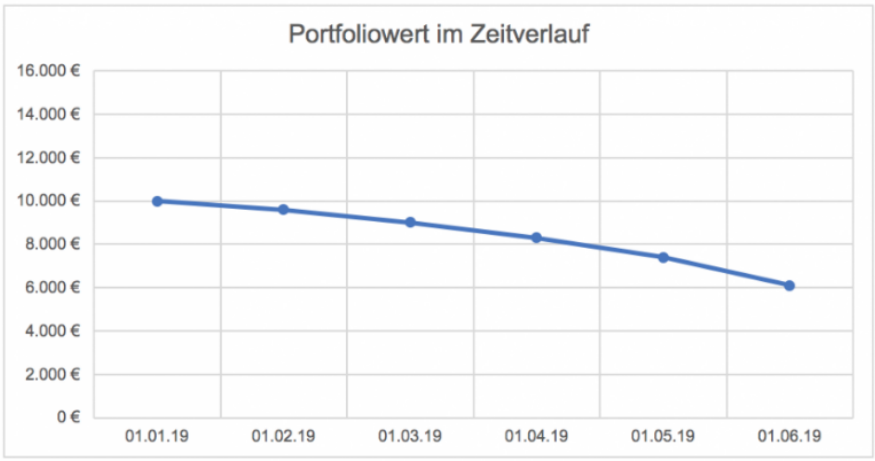
The first question was an experiment where respondents were randomly assigned to one of four groups. The participants then observed a portfolio which was either managed by a) the human advisor named Charles ("Anlageberater Charles"), b) the unnamed human advisor ("Anlageberater"), c) the algorithm named Charles ("Algorithmus Charles") or d) the unnamed algorithm ("Algorithmus").

The information about the portfolio manager was followed by a graphic representation of the portfolio over time. The portfolio was initially worth 10.000€ . The participants were asked at 5 points in time (February until July) whether they wanted to either hold or sell. If they always chose to hold, it looked as follows:

The goal of this question was to find out when participants lose trust in their respective portfolio manager and

[2] Limitations caused by the dependent variable are addressed in Chapter 5. Interpretation, Limitations and Future Research

[3] Limitations caused by the sample will be elaborated in the Chapter 5. Interpretation, Limitations and Future Research



Datum	01.01.19	01.02.19	01.03.19	01.04.19	01.05.19	01.06.19
Gesamtperformance seit Einstieg	0%	-4%	-10%	-17%	-26%	-39%
Wert des Portofolios	10.000 €	9.600 €	9.000 €	8.300 €	7.400 €	6.100 €

Figure 1: Simulated portfolio in the survey

sell. We wanted to conduct an experiment which is realistic. Therefore, it was also possible to never sell the portfolio. This means a cox regression was the way to analyze the data as it is censored (the event of selling does not occur for every participant). When designing the experiment, past research with similar settings was taken into account (Glaser and Walther, 2013). We are convinced that this approach is a good choice since scholars in the past showed that asking for trust directly in self-report measures does not have strong validity and reliability (Cho et al., 2015; Dzindolet et al., 2003; Mcknight et al., 2011).

In the second question, participants had to distribute 10.000€ between an algorithm as an investor and a human advisor. This joint evaluation is a complement to the separate evaluation in question one. Attributes are easier to evaluate when advisors are presented jointly due to increased information (Hodge et al., 2018). The third section was about demographics (age, gender) and if they ever invested money. If they replied no, the survey was over. If they replied yes, they were asked about their investment experience thus far. How much experience did they have? Did they have experience with automated investments? Did they ever use Ginmon? How happy were they with their returns?

4. Results

The survey had 258 respondents (82% male[4], M_{Age} = 36,6, SD_{Age} = 15,7) out of which 233 finished the survey. The sample size for the following analyses varies due to 35 dropouts. Our sample consists of 110 people (70% male, M_{Age} = 27,2, SD_{Age} = 13,5) who have never used the Roboadvisor Ginmon and 113 people (93% male, M_{Age} = 45,6, SD_{Age} = 18,2) who use Ginmon. The sample is not representative for the average German[5]. Due to our sampling method, that was not expected in the first place.

4.1. Experiment I: Simulating a Portfolio with Four Different Portfolio Managers

First, we will look at the results of Experiment I which helps us compare the four different types of portfolio managers.

4.1.1. Behavior of Overall Sample

To get a first overview of the data collected throughout the experiment, the following flow chart shows when and

[4]The impact of the gender distribution on representativeness will be discussed in Chapter 5. Interpretation, Limitations and Future Research.

[5]The limitations for generalization associated with this sampling will be discussed in Chapter 5. Interpretation, Limitations and Future Research.

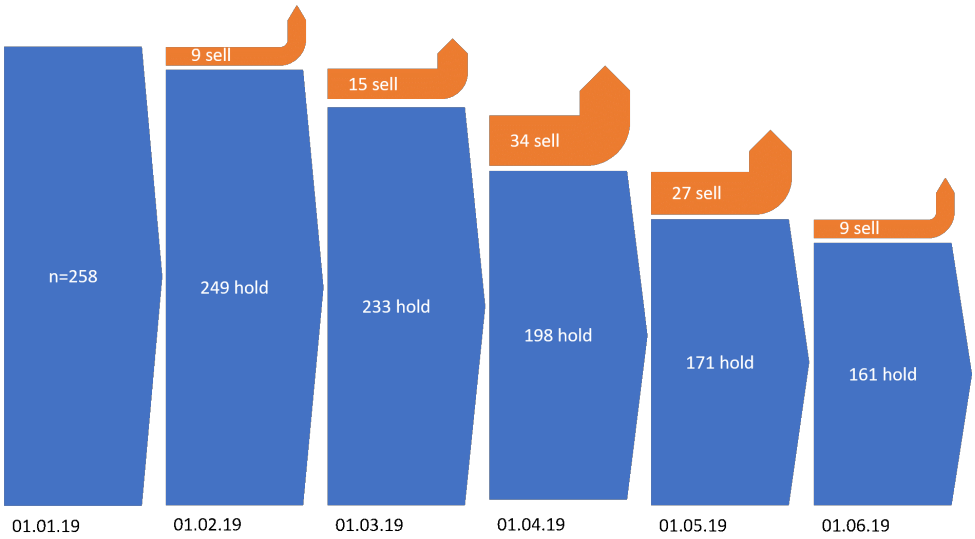


Figure 2: Flow chart depicting sales over time (n = 258)

if participants liquidated their portfolio. It contains the total sample, including all four conditions (Humanized Algorithm, Dehumanized Algorithm, Dehumanized Human and Humanized Human).

As seen in Figure 2, out of 258, 161 (62,4%) never sold their portfolio. 97 respondents (37,6%) sold it. The decline in the portfolio increases over time, nonetheless most people (37) sell in the third month – not in the fifth (9 sales). In general, the respondents held their portfolio for too long. A potential explanation for this phenomenon could be the earlier mentioned loss aversion and sunk costs fallacy. These biases explain why people hold their declining investment portfolios for longer than a rational person should. The respondents are averse to losses which means they tend to not realize their losses by liquidating and "wait it out" instead. In addition, they invested time and energy by tracking the portfolio changes over time. They cannot get the lost time back, so they think it is wasted if they do not manage to sell the portfolio without loss. As a result, they hold the portfolio. Rational behavior would be to only let future costs affect their decisions.

4.1.2. Comparing the Four Portfolio Managers

To analyze the data, we used a cox regression, also called survival analysis. First, we performed a cox regression which compares the experimental groups to the group Humanized Human. We chose the reference group based on Hodge et al. (2018) – we assume this group is the one with the highest loss tolerance and therefore the least sales. Additionally, it makes sense to benchmark the other groups against the Humanized Human since this is still the standard case. When you interact with your bank, you usually talk to a human. Control variables will be added to the model in a later step.

As you can see from Table 3, participants were 2.93 more likely to sell their portfolio when it was managed by a Humanized Algorithm (vs. a Humanized Human) (95% CI, 1.518 - 5.655; P<0.01). When the portfolio was managed by a Dehumanized Algorithm (vs. a Humanized Human), however, participants were (only) around two times (2.093) more likely to sell their portfolio (95% CI, 1.118-3.914; P<0.05). Finally, when the portfolio was managed by a Dehumanized Human (vs. Humanized Human), participants are still around 1.3 times more likely to sell (95% CI, 0.701-2.481; P>0.05[6]).

If we create a ranking of the four groups for "loss tolerance" from highest (least sales) to lowest (most sales) and compare them to findings of Hodge et al. (2018)[7], it looks as follows:

In Table 5, we summarized the results of our Hypotheses 1 to 4: First, in line with H1, we find that participants tolerate losses in their portfolio the most when their portfolio is managed by a Humanized Human. The other three groups have a higher likelihood to sell than the reference group (Humanized Human). Second, our results do not support H2. We assumed respondents' loss tolerance is the second highest for portfolios managed by Dehumanized Algorithms ("Algorithmus") but it ranked third. Third, our results do not support H3. We expected the third highest loss tolerance for the group Humanized Algorithm, but the loss tolerance was the lowest. Fourth, H3 is also not supported by our results. We expected the lowest loss tolerance for the group Dehumanized Human, but loss tolerance was the second highest.

4.1.3. Interaction Effect of Humanizing

H5 and H6 are stating that humanizing has a positive effect (in terms of loss tolerance) for human advisors but a negative effect for algorithmic advisors. Our ranking above sup-

[6]The significance level is discussed in Chapter 4.1.5 Exploratory Analysis.
[7]The differences between both studies regarding the Dependent Variable will be discussed in Chapter 5. Interpretation, Limitations and Future Research.

Table 3: Cox regression (n=258)

Group	B	HR	95% CI		P Values
Humanized Human					
Dehumanized Human	.277	1.319	0.701	2.481	.390
Dehumanized Algorithm	.738	2.093	1.118	3.914	.021
Humanized Algorithm	1.075	2.930	1.518	5.655	.001

Table 4: Comparison of group ranks in two studies

Group	Ranked by loss tolerance (Szeli 2019)	Ranked by likelihood to follow recommendation (Hodge et al., 2018)
Humanized Human	1	1
Dehumanized Human	2	4
Dehumanized Algorithm	3	2
Humanized Algorithm	4	3

Table 5: Results of Hypotheses 1 to 4

Hypotheses	Result
H1: Respondents' loss tolerance is the highest for portfolios managed by Humanized Humans ("Anlageberater Charles").	True
H2: Respondents' loss tolerance is the second highest for portfolios managed by Dehumanized Algorithms ("Algorithmus").	Not true
H3: Respondents' loss tolerance is the third highest for portfolios managed by Humanized Algorithms ("Algorithmus Charles").	Not true
H4: Respondents' loss tolerance is the lowest for portfolios managed by Dehumanized Humans ("Anlageberater").	Not true

ports this reasoning; the Humanized Human (Rank 1) outperforms the Dehumanized Human (Rank 2), but the Humanized Algorithm (Rank 4) underperforms the Dehumanized Algorithm (Rank 3). To test these predictions quantitatively, we computed a 2 (advisor: human vs. algorithm) x 2 (humanization: humanized vs. dehumanized) cox regression model (Table 6).

The main effect of the type of advisor is marginally significant (p=.064). Given the b-value of -.461 and the coding of our variables, we can say that portfolios managed by Human Advisor were sold less than portfolios managed by an Algorithmic Advisor. The interaction effect between the factors humanization (named vs. unnamed) and advisor (human vs. algorithm) was not significant (p=.133). That is, the effect of humanization on selling is not statistically different for a human (vs. algorithm). According to the ranking reported in Table 4 as well as our Hypotheses 5 and 6, we assumed a different finding. We expected a significant interaction effect since the Humanized Human had the highest loss tolerance while the Humanized Algorithm had the lowest loss tolerance (Table 4). Thus, we will take a closer look at the changes in significance in later chapters, if we split the data set (4.1.5 Comparing users and non-users) or increase sample size (4.1.6 Exploratory analysis).

4.1.4. Prior Investment Experiences and Demographics

In this chapter, we analyze variables related to prior investment experiences and demographics. 31 respondents (14%) have never invested money. The remaining 190 (86) consider their experience with investing money and automated investments as follows:

Most subjects consider themselves "advanced" investors, only a few edge cases are "newbies" or "experts".

In automated investing, more subjects are newbies and beginners:

The majority (Figure 5) had at least neutral experiences with the returns on their investment in the past. Only a few are very unhappy. One possible explanation for that can be that many respondents are in their twenties which means they have been investing in a bull market (since 2009) all their adult life.

Earlier studies have shown that demographic factors and past investment experiences play an important role in people's future investment behavior (Grinblatt and Keloharju, 2000; Schnell, 2011; Statman, 1999). Therefore, we decided to calculate an extension of the initial cox regression model (Table 3) by controlling for the following factors: Age, Gender, Satisfaction with past returns, Experience with investing

Table 6: Main effects and interaction effects (Cox Regression; n=256)

Effect	df	B	SE	HR	P Values
Humanization (named/unnamed)	1	.277	.322	1.319	.390
Advisor (Algorithm/Human)	1	-.461	.249	.630	.064
Humanization x Advisor	1	-.614	.414	.541	.138

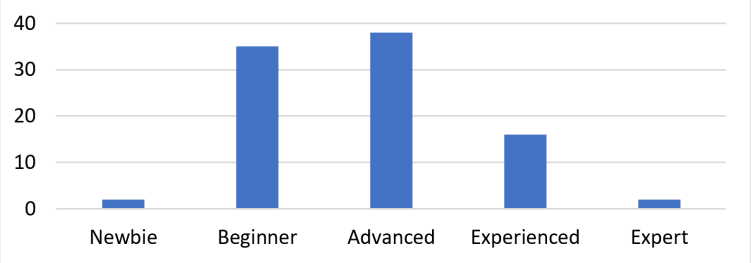


Figure 3: Experience with investing money (n=190)

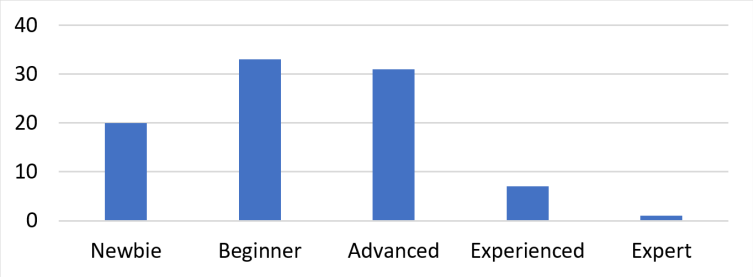


Figure 4: Experience with automated investments (n=190)

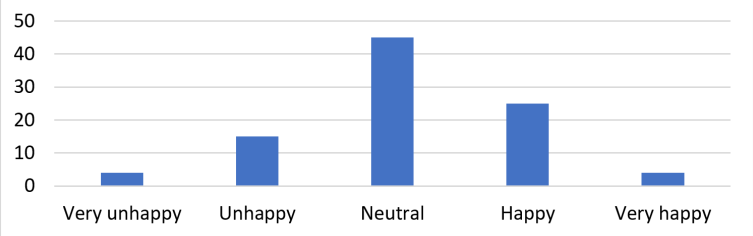


Figure 5: Satisfaction with past returns

and Experience with automated investing (see Table 7).

Table 7 shows that all of the added factors are not significant. Moreover, we find that the coefficient of the four different portfolio managers do not change substantially, if we include this set of control variables. Thus, these findings suggest that the results reported in 4.1.2 hold true controlling for demographical variables (e.g. gender, age) and for previous financial experiences (e.g. experiences with automated investing).

4.1.5. Comparing Users and Non-users

In this chapter, we compare Ginmon users (N = 110, 93% male, M_{Age} =45,6, SD_{Age} = 18,2) versus Ginmon non-users (who have never used any other Roboadviser) (N = 113, 70% male, M_{Age} =27,2, SD_{Age} = 13,5)[8].

First, we ran the cox regression (Table 3) again – but this time with a split dataset. Table 8 reports differences between users and non-users.

As one can see, results across Ginmon users versus non-users (Table 8) are very similar to the results without splitting the data (Table 3), since the ranking of the groups relative to the reference group Humanized Human is the same.

If we compare users to non-users, particularly striking is that the p-values for all groups are lower for non-users than for users while all the b-values are higher. That is, users are

[8]The differences between in age and gender will be discussed in Chapter 5. Interpretation, Limitations and Future Research.

Table 7: Extended Cox regression (n=258)

Group	B	HR	95% CI		P Values
Humanized Human					
Dehumanized Human	.098	1.104	0.535	2.276	.790
Dehumanized Algorithm	.651	1.918	0.943	3.899	.072
Humanized Algorithm	1.020	2.773	1.304	5.897	.008
Age	-.004	.996	.982	1.010	.575
Gender	-.440	.644	.341	1.217	.175
Satisfaction with past returns	-.191	.826	.621	1.101	.192
Experience with investing	.176	1.192	.867	1.639	.280
Experience with automated investing	.029	1.030	.763	1.389	.849

Table 8: Cox regression for users (n=113) and non-users (n=110)

Cox regression for users (n=113)					
Group	B	HR	95% CI		P Values
Humanized Human					
Dehumanized Human	.384	.681	.242	1.914	.466
Dehumanized Algorithm	.826	.439	.160	1.203	.110
Humanized Algorithm	.916	.400	.143	1.122	.082
Cox regression for non-users (n=110)					
Humanized Human					
Dehumanized Human	.475	.622	.313	1.237	.176
Dehumanized Algorithm	.823	.439	.216	.892	.023
Humanized Algorithm	1.531	.216	.098	.477	.000

less likely to sell their portfolio across all four conditions. Compared to the original sample (n=256), for non-users we found lower p-values despite reduced sample size (n=110)[9].

Second, we ran the cox regression with a 2x2 design testing for main and interaction effects (like in Table 6). But this time, the analysis was split by users and non-users (Table 9).

For non-users, the interaction effect (p=.015) is significant. Thus, for nonusers, humanizing has a different effect on the time of selling the portfolio depending on whether the advisor is a human or an algorithm. Humanizing the human leads to less sales while humanizing the algorithm leads to more sales (compared to the non-humanized counterpart). For users, however, we cannot make the same statement since the effect of humanizing was not significant (p=.454). Given Table 9 and the results of our ranking, we can say that the following hypotheses hold true for nonusers:

H5: Respondents' loss tolerance is higher for portfolios managed by Humanized Human ("Anlageberater Charles") than for a Dehumanized Human.

H6: Respondents' loss tolerance is higher for portfolios managed by Dehumanized Algorithm ("Algorithmus") than for a Humanized Algorithm ("Algorithmus Charles").

4.1.6. Exploratory Analysis

The interaction effect in the overall sample (Table 6) as well as in the non-user group (Table 9) was not significant. To explore whether and how this would change with increased sample size, we (artificially) duplicated our dataset (Table 10).

Given the larger sample size, the interaction effect (p=.036) and the main effect of the advisor type (p=.009) are significant. Based on the b-value (-.461) and our coding, we can say that there is a main effect which means that portfolios managed by a human were less likely to be sold than portfolios managed by the algorithm. Based on these findings, one could speculate that rerunning the experiment — with an increased sample size — could yield a significant interaction effects of the factor Advisor and Humanization x Advisor (and therefore H5 and H6 would hold true)[10]. There needs to be additional empirical validation in the future. We suggest a bigger sample size or a research design which allows to test with an ANOVA, so significant effects can be found more easily due to uncensored data.

Since none of the control variables were significant, we also decided to rerun a cox regression including factors which

[9] Possible explanations for the differences between users and non-users and ideas for future research are reported in Chapter 5. Interpretation, Limitations and Future Research.

[10] More details on potential future research can be found in Chapter 5. Interpretation, Limitations and Future Research.

Table 9: Main effects and interaction effects for users (Cox Regression; n=113) and non-users (n=110)

Main effects and interaction effects for users (Cox Regression; n=113)					
Effect	df	B	SE	HR	P Values
Humanization (named/unnamed)	1	.385	.528	1.469	.466
Advisor (Algorithm/Human)	1	-.370	.370	.645	.235
Humanization x Advisor	1	-.477	.636	.621	.454
Main effects and interaction effects for non-users (Cox Regression; n=110)					
Humanization (named/unnamed)	1	.475	.351	1.608	.176
Advisor (Algorithm/Human)	1	-.348	.288	.706	.226
Humanization x Advisor	1	-1.183	.488	.306	.015

Table 10: Main effects and interaction effects with doubled sample size (Cox Regression; n=512)

Effect	df	B	SE	Exp(B)	P Values
Humanization (named/unnamed)	1	.277	.322	1.319	.224
Advisor (Algorithm/Human)	1	-.461	.249	.630	.009
Humanization x Advisor	1	-.614	.414	.541	.036

were closest to marginal significance level. Again, we doubled the sample size (Table 11) to explore, if our sample size is the reason the control variables are not significant.

Interestingly, gender is now significant (p=.04). Also, prior investment experience (p=.108) and satisfaction with returns (p=.078) have p-values close to the marginal significance level. Given the negative b-values of Gender (B = -.445; Coding: 1 = female, 2 = male) and Satisfaction with returns (B = -.180; Coding: 1 = very unhappy, 5 = very happy), we can conclude that females and respondents who were unhappy with the past performance of their portfolio were more likely to sell than their counterparts. The positive b-value of investment experience (B = .163; Coding: 1 = Newbie, 5 = Expert), means that more experienced people were more likely to sell than less experienced people. This supports our findings where we compared users to non-users. As stated earlier, these results should encourage future researchers to run an additional study with larger sample size.

4.2. Experiment II: Allocation of the Budget Between Human and Algorithm

In one question we asked in the survey, respondents had a choice between the human and the algorithm (allocating an investment of 10.000€ between both of them). The following pie charts represent the allocation of the 10.000€ between the two advisors. The first interesting finding is that most (70%) of the money was given to the algorithm.

This is an extreme deviation from the actual distribution of investments managed by human advisors versus algorithmic advisors (7,7 Million EUR in 2019 (Statista, 2018)) in Germany. This is also an interesting finding in the light of the results of Experiment I. Respondents had a higher loss tolerance with human advisors but still claim that they would allocate most of their money to an algorithm[11].

Second, we distinguish between people who do not use Ginmon and people who use Ginmon. There is substantial difference: Ginmon users allocate a lot more money (80%) to the algorithmic advisor than non-users (59%). People with more experience in algorithmic were willing to give more money to the algorithm. A Z-proportions test supports this difference (Z = 2.868, p = 0.002).

5. Interpretation, Limitations and Future Research

This chapter will cover interpretation of our results, limitations of this research and suggestions for future researchers.

5.1. Experiment I: Simulating a Portfolio with Four Different Portfolio Managers

Some of our findings (H1, H5, H6) were coherent with Hodge et al. (2018), while others were not (H2, H3, H4). Regarding the coherent findings in H1, it remains to be said that Humanized Humans are still the most trusted decision-makers with the highest loss tolerance among investors. This finding cumbers the mass adoption for Roboadvisors, as they will always be benchmarked against the human equivalent. As a result, many companies started to humanize their advisors – probably as an attempt to gain some of the trustworthiness that is attributed to Humanized Humans. As H5 and H6 show, doing that can have the opposite of the hoped effect –

[11] Possible explanations for the observed differences in budget allocation (human versus algorithm) between in Experiment II, real world behavior and Experiment II are reported in Chapter 5. Interpretation, Limitations and Future Research.

Table 11: Extended Cox regression (n=516)

Group	B	HR	95% CI		P Values
Humanized Human					
Dehumanized Human	.093	1.097	0.659	1.829	.721
Dehumanized Algorithm	.627	1.872	1.140	3.074	.013
Humanized Algorithm	.982	2.669	1.586	3.490	.000
Gender	-.445	.641	.419	.981	.040
Satisfaction with past returns	-.180	.835	.684	1.020	.078
Experience with investing	.163	1.117	.965	1.436	.108

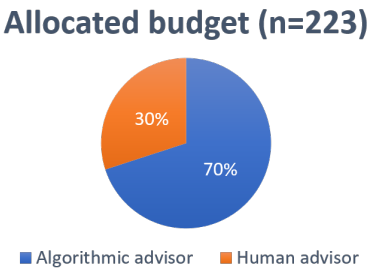


Figure 6: Allocation of 10.000€ (n=223)

Figure 7: Allocation of 10.000€ split in non-users (n=110) and users of Ginmon (n=113)

at least for people who have never used a Roboadvisor. Humans seem to trust algorithms for their non-human traits (rational, neutral, high processing power) and are not looking for the same traits that are attributed to humans (empathy, accountability). When interpreting the results of H5 and H6, it makes sense to call oneself the following in mind: A technology can be perceived simple and unable to complete complex tasks, if it is named (Hafer et al., 1996; Riegelsberger et al., 2005). For human advisors, the trust level is increased, if they share personal information, as it demonstrates their willingness to risk their reputation (Garner, 2005; O'Keefe, 1990; Pornpitakpan, 2004). Note that a high trust level in itself is neither positive nor negative, but neutral. For an investor, it can also be beneficial to lose trust early, as we have seen in our case with the declining portfolio.

5.2. Differences to Hodge et al. (2018)

Regarding H2, H3 and H4, we have to say that ranks two to four are not coherent with Hodge et al. (2018) findings. Especially, Dehumanized Humans ranked higher in our study. When comparing both studies, we have to keep in mind that the dependent variables were not the same. We measured, if and when they sell their portfolio, while the other study asked for their likelihood to follow the investment advice. Another possible explanation for the deviations: In our experiment the advisor was also the one making the investment decision, whereas in Hodge's experiment it was just an investment recommendation. In definitive and binding investment decision-making people seem to still prefer humans over algorithms. However, when it is just about recommendations, they might prefer an algorithm over a human. Future research should look deeper into that thesis. In addition to that, the sample (MBA students versus users and non-users of Roboadvisors) could be a reason for the diverging results. Furthermore, Hodge provided extensive background information while in our case the subjects had to "trust blindly". Most of the existing research focuses on AI as a recommendation engine which requires much lower trust levels than in settings with an algorithm as the final decision-maker. There-

fore, we encourage other scholars to study experimental settings which allocate more power to the algorithm. A future study could test, if a higher amount of background information about the investment decision leads to higher trust in algorithmic decisions. This would help exploring, if the differences in ranking placements two to four are caused by the diverging background information provided.

5.3. Experiment II: Allocation of the Budget Between Human and Algorithm

Respondents said they would allocate 70% of their budget to the algorithm, even though in reality they behave differently. There seems to be a difference between people's intentions and their actions. Experiment I has shown that the portfolios managed by the algorithm are sold earlier, while in Experiment II, respondents claim to trust the algorithm more than the human. One reason could be social desirability. Maybe some respondents think trusting AI is what the researchers expect as a result (Rosenthal effect). Especially, since parts of the sample were recruited using the author's personal social media account.

The result of 70% of budget allocated to the algorithm is also an extreme deviation from the real world: In 2019, Roboadvisors had only 7,7 Million EUR under management in Germany (Statista, 2018). A possible explanation for the difference could be certain patterns in our sample (young, many users of Roboadvisors) which prevent generalizations. Another reason could be that we forced respondents to compare between two options, which is not a realistic scenario. Future research could dig deeper into the divergence between intentions (allocating money to the algorithm) and actions (trusting humans more).

Additionally, it needs to be mentioned that the question from which this data originates was asked after Experiment I, which simulated a declining portfolio. Depending on who managed the declining portfolio, it might affect the answers to the subsequent question. Since our respondents are randomly assigned to groups with equal sizes, this effect can be slightly reduced.

5.4. Limitations Caused by the Sample

The overall sample (N = 223, 82% male, M_{Age} = 36,6, SD_{Age} = 15,7) is not representative for the "average German" for three reasons.

Firstly, our sample is too young since the non-user group (N = 110, 70% male, M_{Age} = 27,2, SD_{Age} = 13,5) mainly consisted of students. The median age of Germans is 44,3 years (UN, 2013). The age difference should not be neglected, since age is a factor in trust in AI and digital products in general (Scopelliti et al., 2005; Evers et al., 2008; Ho et al., 2005; Hancock et al., 2011). Secondly, our sample is male-dominated (82%). This might be representative for users of Roboadvisors or investors in general (Schnell, 2011), but not for the general population. If the goal is to find out how Roboadvisor could reach broader adoption, it is recommended to also look at potential users which are underrepresented in the current user base (e.g. females). Thirdly, users were overrepresented in our sample compared to the German population. Since this is an exploratory work, the goal was not to have a representative sample for "the average German". Nonetheless, the limitations caused need to be mentioned.

An additional limitation comes with the limited sample size. The explorative analyses showed that a larger sample size could be beneficial. For example, a significant p-value for the Dehumanized Human group in the cox regression (Table 3) might be found with an increased sample size.

5.5. Comparing Users to Non-users

When we compare the two subgroups of users and non-users, we need to keep in mind that the non-user group is a lot younger and contains less females. These factors could be intervening variables when we try to make statements about the effects of being a user of a Roboadvisor.

When opposing users to non-users, the main conclusion is that we found lower p-values and higher b-values for the cox regression comparing all four advisor types (Table 8) as well as for the cox regression with a 2x2 design (Table 9).

An explanation could be that the Ginmon users are more used to portfolio volatility – no matter who manages it. They are already familiar with a somewhat humanized algorithm trading for them (Ginmon's algorithm is named Apeiron). They have more experience and therefore were harder to manipulate with our experimental stimulus. In contrast, non-users, who most likely interact with a humanized algorithm for the first time, might be more prone to manipulation. Consequently, the effect of humanization and advisor type is weakened compared to the non-users. You could say users are more immune to the effect of humanization.

In other words, to investigate effects of humanization we recommend future researchers taking a closer look at non-users. You could choose to observe people who have no experience with Roboadvisors to gain a deeper understanding of the differences between the influence of the four portfolio managers on loss tolerance. Also, you could gain a deeper understanding of diverging effect of humanizing humans versus algorithms.

5.6. Further Limitations and Future Research

Given the results of our exploratory analysis (increased sample size, control varaibles) and user versus non-user comparison, we can conclude that future researcher should try to recruit a bigger sample size, incorporate control variables and focus on non-users without prior automated investment experiences.

Trust in Artificial Intelligence is a rather broad topic. For the sake of feasibility, we had to focus on one specific manifestation of trust (loss tolerance) and one application of AI (Roboadvisors).

Many researchers tried to operationalize trust but did not arrive at an overarching conclusion. Therefore, for the dependent variable we decided to focus on one key aspect which made sense in the context of investing: How do investors react to losses? Do they sell or do they hold? This

comes with the limitation that we did not directly operationalize and measure trust. Future research could try to look at the bigger picture of trust, e.g. by surveying multiple items which cover more than loss reaction and loss tolerance. The study can be based on other scholars' attempts to quantify trust. For example, Cho et al. (2015) developed a scale ranging from complete distrust to complete trust including undistrust and untrust (Cho et al., 2015).

Also, our choice of the independent variable leads to certain shortcomings. As mentioned in the literature review, there are dozens of factors which potentially influence user trust in AI. In order to not exceed the limits of this research, we manipulated just one variable: humanization. Humanization is of great interest, as many companies in the space name their technologies or represent them humanlike (avatars, chatbots) – especially in finance. Since the effect on users is controversial in research, we decided to contribute to that discussion. Within the factor humanization, we – again – had to narrow it down. We named the decision-maker, but you could also humanize in other ways, e.g. give him/her a face or a voice. Future research could look at other factors which could influence user trust and at other interesting ways of humanization.

Another limitation is caused by the scope of our study. We cannot make general statements about AI as a whole. The domain is too broad due to the manifold of applications in different industries. We had to focus on AI-based finance products, more specifically Roboadvisors. Humanizing decision-makers could have the opposite effect (to what we found) in other domains such as entertainment, navigation or healthcare. Conducting a study in other domains is recommended. In general, due to the context-dependency of trust, studies in non-finance environments are also beneficial.

In our survey, we simulated the development of a portfolio over time. Obviously, the ideal and more realistic study would be a longitudinal study with real money to lose. Due to time- and budget-constraints this was not feasible, but maybe it is for other scholars. To give a concrete recommendation: We would conduct a similar survey, just in the mentioned realistic setting. Instead of just asking, if the subjects would like to hold or sell, we would add more questions on trust. For example, you could ask a question such as: Do you consider the decision-maker trustworthy? You could even offer them to switch to another decision-maker (human/algorithm) to allow for joint evaluation. If respondents answer the questions monthly, and not just directly one after another, it could lead to interesting insights and we could learn more about the change in trust over time.

Additionally, the simulation of the declining portfolio as the first part of the survey biases the questions afterwards. We still decided to put it first since the question was the core of our hypotheses and we wanted to maximize the sample size for the cox regression. Putting other questions first would have distorted subsequent answers as well. The bias should be bearable, as the portfolio development was identical across all four experimental groups. Nonetheless, future studies could run separate surveys to prevent that shortcoming.

6. Implications

Humanizing AI seems not to be a sufficient solution to increase users trust in AI – at least in the finance domain. Also, the trust in AI (in terms of allocated investments) differs across demographics. You can ask yourself, if increasing the trust level should always be the end goal. In our experiment, as in many investment decisions in real life, you were better off (financially), if you lost trust sooner. The findings of this work have several implications on three different levels: Regulation, User Experience and Technology.

Regulation is an important topic, given the observed willingness of some consumers to allocate a substantial amount of money to algorithms. But regulators were not able to keep up with the rapid technological progress. Regulating AI in general is most likely not suitable since there are plenty of different applications and domains. Nonetheless, there is consensus that a global ethical standard for AI companies would be beneficial. The challenge is to protect fundamental rights and freedom while still encourage innovation of AI technologies. Especially, in our context of humanization and investments, there are issues arising when the borders between human and machine become blurred. Formal advancements by the EU are concerned with anti-discrimination, biases and fairness of algorithms. Also, in the finance domain, this can be an issue (e.g. credit scoring). Thelisson (2017) proposes four solutions: First, a Code of Conduct that one or multiple companies bind themselves to. Second, Quality Labels and Audits. This means the company either discloses the code publicly (issues with interpretability and IP remain) or hires another authority that looks into the code and certifies it with some kind of quality seal. Third, transparency in the data chain. And fourth, de-biasing datasets and algorithms which can be broken down to discrimination detection and discrimination prevention (Thelisson et al., 2017). The GDPR constitutes that individuals have the right to object to decisions even if they are made purely on the basis of automation. In addition, they also have the right to obtain information about the existence of an automated decision making system as well as the "logic involved" and its consequences (Thelisson, 2017). How this right of explanation plays out practically (e.g. depth of explanation, technical feasibility) will be decided in court in the future. The technological aspects of the explainability will be analyzed in the last paragraph of this chapter.

User experience also plays an important role to build trust in AI-based finance products. While humanization was not beneficial in our case, it still needs to be explored for other products or to other degrees (e.g. avatars). Other scholars found out that users prefer conversational agents which are young, match their ethnicity and show non-verbal cues while gender does not matter (Cowell and Stanney, 2003). These agents do not only humanize the product but also increase personalization and familiarity. There is empirical evidence

that these factors increase trust and product adoption (Komiak and Benbasat, 2006). From a UX perspective, it could also be beneficial to give users the option to interact with and try the algorithm. Dietvorst et al. (2018) showed that giving users the opportunity to modify the algorithm reduces algorithm aversion. Giving a share of control back to the human may sound like a contrast to the idea of a Roboadvisor (full automation), but even slight modifications increase the users' preference for that algorithm (Dietvorst et al., 2018). Giving the users the possibility to make optional slight alterations (e.g. changing the risk class after the initial investment) could be a solution. Another factor is accountability. Who is responsible if the algorithm loses your money? This can be considered as UX since the product needs to convey a sense of security (e.g. to make up for lack of accountability). One solution for startups, which do not have a well-known brand yet, is to offer the Roboadvisor as a whitelabel solution of an established bank. But the accountability issue also involves technological aspects (error-proneness) and legal aspects (Who can be hold accountable by law?). Additionally, UI-decisions, such as design, should also have the goal to increase the Roboadvisors' trustworthiness. An option to overcome the lack of trust in Roboadvisors is to slowly move customers from a conventional bank with a human advisor to a Roboadvisor. This allows them to downsize advisory operations. High net-worth individuals can still have their human account while others can also have the option to take part in economic growth by investing without expert knowledge. This is especially crucial for risk-averse, inexperienced, low-budget individuals (Jung et al., 2018).

Humanization is not a technological challenge in itself and also not beneficial according to the conducted study. Therefore, the technological implications evolve mainly around transparency of algorithms. Many users would be interested in understanding how the algorithm arrives at its conclusions by explaining its reasoning. For example, if your Roboadvisor loses money, you would like to know how it happened and seek a justification for the action that lead to losses (e.g. a specific trade) (Gregor and Benbasat, 1999). This kind of explainable AI would make decision-making more transparent and predictable and therefore easier to trust.

7. Conclusion

To conclude, we can say that we arrived at three main findings. Firstly, for non-users humanizing a human decision-maker increases loss tolerance, while humanizing an algorithm decreases loss tolerance. Given the slight manipulation (difference of one word), this finding is astonishing. Secondly, the humanized Human is still the most trusted financial advisor. Thirdly, despite the second finding: Respondents were willing to allocate much more money to the algorithmic advisor compared to the human advisor. This leaves us with many opportunities for future research since the scope and therefore potential generalization was limited in our study.

On a theoretical level, the conclusion is that there are many theories concerning trust in AI but most of them focus on automated recommendations for action and the human as the final decision maker. A structured, comprehensive model that includes the high number of trust factors and the algorithm as the decision maker is missing. With the current raise in automation and the complexity of trust, researchers from different disciplines shall collaborate to build a holistic model.

The question determining the next decades is not about technological feasibility. The amount of data available grows every day. The crucial question is, if users will trust in the decisions made by the algorithms. Given the current levels of users trust in Roboadvisors, a hybrid model of virtual advice in wealth management that combines human and algorithm seems to be realistic (Cocca, 2016).

References

Adams, B., Bruyn, L., Houde, S., and Angelopoulos, P. Trust in automated systems literature review. Defense research and development Canada Toronto No. Technical report, CR-2003-096, 2003.

André, E., Gimpel, H., and Olenberger, C. An investigation of the effects of anthropomorphism in collective human-machine decision-making. Technical report, 2018.

Arkes, H. R. and Blumer, C. The psychology of sunk cost. *Organizational Behavior and Human Decision Processes*, 35(1):124–140, February 1985.

Bernstein, P. L. *Against the gods: the remarkable story of risk*. New York, John Wiley & Sons, 1996.

Bottazzi, L., Da Rin, M., and Hellmann, T. The importance of trust for investment: evidence from venture capital. *The Review of Financial Studies*, 29(9):2283–2318, 2016.

Burgoon, J. K., Bonito, J. A., Bengtsson, B., Cederberg, C., Lundeberg, M., and Allspach, L. Interactivity in human–computer interaction: a study of credibility, understanding, and influence. *Computers in Human Behavior*, 16(6):553–574, November 2000.

Carré, O., James, F., and Witz, P. Keynote presentation automation in financial advice: Trends, opportunities and implications. *The Luxembourg Bankers' Association event Robo-Advisors and Digital Investments*, 2016.

Castelfranchi, C. and Falcone, R. *Trust theory: a socio-cognitive and computational model*, volume 18. Chichester, John Wiley & Sons, 2010.

Chaminade, T., Hodgins, J., and Kawato, M. Anthropomorphism influences perception of computer-animated characters' actions. *Social Cognitive and Affective Neuroscience*, 2(3):206–216, September 2007.

Chin, S. H. On application of game theory for understanding trust in networks. In *2009 International Symposium on Collaborative Technologies and Systems*, pages 106–110. IEEE, 2009.

Cho, J., Swami, A., and Chen, I. A survey on trust management in mobile ad hoc networks. *IEEE Communications Surveys Tutorials*, 13(4):562–583, Fourth 2011. ISSN 2373-745X. doi: 10.1109/SURV.2011.092110.00088.

Cho, J.-H., Chan, K., and Adali, S. A survey on trust modeling. *ACM Computing Surveys (CSUR)*, 48(2):1–40, 2015.

Cocca, T. D. Potential and limitations of virtual advice in wealth management. *The Capco Institute Journal of Financial Transformation*, 44:45–57, 2016.

Coeckelbergh, M. Can we trust robots? *Ethics and information technology*, 14(1):53–60, 2012.

Cowell, A. J. and Stanney, K. M. Embodiment and interaction guidelines for designing credible, trustworthy embodied conversational agents. In *International workshop on intelligent virtual agents*, pages 301–309. Springer, 2003.

Dawes, R. M. The robust beauty of improper linear models in decision making. *American psychologist*, 34(7):571, 1979.

Diab, D. L., Pui, S.-Y., Yankelevich, M., and Highhouse, S. Lay perceptions of selection decision aids in us and non-us samples. *International Journal of Selection and Assessment*, 19(2):209–216, 2011.

Dietvorst, B. J., Simmons, J. P., and Massey, C. Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1):114, 2015.

Dietvorst, B. J., Simmons, J. P., and Massey, C. Overcoming algorithm aversion: people will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, 64(3):1155–1170, 2018.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., and Beck, H. P. The role of trust in automation reliance. *International journal of human-computer studies*, 58(6):697–718, 2003.

Eastwood, J., Snook, B., and Luther, K. What people want from their professionals: attitudes toward decision-making strategies. *Journal of Behavioral Decision Making*, 25(5):458–468, 2012.

Einhorn, H. J. and Hogarth, R. M. Decision making under ambiguity: a note. In *Risk, decision and rationality*, pages 327–336. Springer, 1988.

Epley, N., Waytz, A., Akalis, S., and Cacioppo, J. T. When we need a human: Motivational determinants of anthropomorphism. *Social Cognition*, 26 (2):143–155, April 2008.

Epperson, T., Hedges, B., Singh, U., and Gabel, M. Hype vs. reality: The coming waves of "robo" adoption. Technical report, A.T. Kearney, 2015.

Eule, A. Rating the robo-advisors. barron's-the dow jones business and financial weekly, 2017. URL https://www.barrons.com/articles/rating-the-robo-advisors-1501303316.

Evers, V., Maldonado, H., Brodecki, T., and Hinds, P. Relational vs. group self-construal: untangling the role of national culture in hri. In *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 255–262. IEEE, 2008.

Eyssel, F., Ruiter, L. D., Kuchenbrandt, D., Bobinger, S., and Hegel, F. 'if you sound like me, you must be more human': On the interplay of robot and user features on human-robot acceptance and anthropomorphism. URL https://aiweb.techfak.uni-bielefeld.de/files/p125.pdf.

Gambetta, D. 'Can We Trust Trust?', in Gambetta, Diego (ed.) Trust: Making and Breaking Cooperative Relations, electronic edition, Department of Sociology, University of Oxford, chapter 13, pp. 213-237, 2000. URL http://www.sociology.ox.ac.uk/papers/gambetta213-237.pdf.

Garner, R. What's in a Name? Persuasion Perhaps. *Journal of Consumer Psychology*, 15(2):108–116, 2005.

Glaser, M. and Walther, T. Run, Walk, or Buy? Financial Literacy, Dual-Process Theory, and Investment Behavior. *SSRN Electronic Journal*, 2013.

Gold, N. A. and Kursh, S. R. Counterrevolutionaries in the financial services industry: Teaching disruption-a case study of roboadvisors and incumbent responses. *Business Education Innovation Journal*, 9(1):139–146, 2017.

Gong, L. How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior*, 24(4):1494–1509, July 2008.

Gregor, S. and Benbasat, I. Explanations from intelligent systems: Theoretical foundations and implications for practice. *MIS Quarterly*, 23(4):497, December 1999.

Greiner, M. E. and Wang, H. Building consumer-to-consumer trust in e-finance marketplaces: An empirical analysis. *International Journal of Electronic Commerce*, 15(2):105–136, 2010.

Grinblatt, M. and Keloharju, M. The investment behavior and performance of various investor types: a study of finland's unique data set. *Journal of Financial Economics*, 55(1):43–67, January 2000.

Grove, W. M. and Meehl, P. E. Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, public policy, and law*, 2(2):293, 1996.

Guiso, L., Sapienza, P., and Zingales, L. Trusting the stock market. *The Journal of Finance*, 63(6):2557–2600, December 2008.

Guthrie, S. E. *Faces in the Clouds: A New Theory of Religion*. Oxford, Oxford University Press, 1993.

Hafer, C. L., Reynolds, K. L., and Obertynski, M. A. Message comprehensibility and persuasion: Effects of complex language in counterattitudinal appeals to laypeople. *Social Cognition*, 14(4):317–337, December 1996.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., and Parasuraman, R. A meta-analysis of factors affecting trust in human-robot interaction. *Human factors*, 53(5):517–527, 2011.

Haugen, R. A. *The new finance: the case against efficient markets*. Henglewood Cliffs, Prentice Hall, 1995.

Highhouse, S. Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology*, 1(3):333–342, 2008.

Ho, G., Wheatley, D., and Scialfa, C. T. Age differences in trust and reliance of a medication management system. *Interacting with Computers*, 17(6): 690–710, 2005.

Hodge, F. D., Mendoza, K. I., and Sinha, R. K. The effect of humanizing robo-advisors on investor judgments. *Available at SSRN 3158004*, 2018.

Houser, D., Schunk, D., and Winter, J. Distinguishing trust from risk: an anatomy of the investment game. *Journal of Economic Behavior & Organization*, 74(1-2):72–81, May 2010.

James, H. S. The trust paradox: a survey of economic inquiries into the nature of trust and trustworthiness. *Journal of Economic Behavior & Organization*, 47(3):291 – 307, 2002. ISSN 0167-2681. doi: https://doi.org/10.1016/S0167-2681(01)00214-1. URL http://www.sciencedirect.com/science/article/pii/S0167268101002141.

Jung, D., Dorner, V., Weinhardt, C., and Pusmaz, H. Designing a robo-advisor for risk-averse, low-budget consumers. *Electronic Markets*, 28(3):367–380, August 2018.

Kahneman, D. and Tversky', A. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292, 1979.

Komiak, S. Y. X. and Benbasat, I. The effects of personalization and familiarity on trust and adoption of recommendation. Technical report, 2006.

Krause, P. and Schäfer, A. Verteilung von Vermögen und Einkommen in Deutschland: Große Unterschiede nach Geschlecht und Alter. *DIW*

Wochenbericht, 72(11):199–207, 2005.

Kydd, A. H. *A blind spot of philosophy. Trust and Mistrust in International Relations*. Princeton University Press, 2005.

Lagerspetz, O. *Trust: the tacit demand*, volume 1. Luxembur, Springer Science & Business Media, 1998.

Leaden, R. 9 of the Funniest and Most Shocking AI Fails, 2017. URL https://www.entrepreneur.com/slideshow/289621.

Lee, J. D. and See, K. A. Trust in automation: designing for appropriate reliance. *Human factors*, 46(1):50–80, 2004.

Lee, S., Choi, J., Ngo-Ye, T. L., and Cummings, M. Modeling trust in the adoption decision process of robo-advisors: an agent-based simulation approach. In *Proceedings of the Tenth Midwest Association for Information Systems Conference, Saint Louis, MO, USA*, pages 17–18, 2018.

Logg, J. M., Minson, J. A., and Moore, D. A. Algorithm appreciation: people prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151:90–103, 2019.

Mayer, R. C., Davis, J. H., and Schoorman, F. D. An integrative model of organizational trust. *Academy of management review*, 20(3):709–734, 1995.

Mcknight, D. H., Carter, M., Thatcher, J. B., and Clay, P. F. Trust in a specific technology: an investigation of its components and measures. *ACM Transactions on management information systems (TMIS)*, 2(2):1–25, 2011.

Millward, P. The 'grey digital divide': perception, exclusion and barriers of access to the internet for older people. *First Monday*, 8(7), July 2003.

Nan, X., Anghelcev, G., Myers, J. R., Sar, S., and Faber, R. What if a Web Site can Talk? Exploring the Persuasive Effects of Web-Based Anthropomorphic Agents. *Journalism & Mass Communication Quarterly*, 83(3): 615–631, September 2006.

Odean, T. Are investors reluctant to realize their losses? *The Journal of Finance*, 53(5):1775–1798, October 1998.

O'Keefe, D. J. *Persuasion: theory and research*. Thousand Oaks, SAGE Publications, 1990.

Önkal, D., Goodwin, P., Thomson, M., Gönül, S., and Pollock, A. The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making*, 22(4):390–409, 2009.

Payne, J. W. Relation of perceived risk to preferences among gambles. *Journal of Experimental Psychology: Human Perception and Performance*, 1(1): 86–94, 1975.

Pornpitakpan, C. The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology*, 34(2): 243–281, February 2004.

Promberger, M. and Baron, J. Do patients trust computers? *Journal of Behavioral Decision Making*, 19(5):455–468, 2006.

Riegelsberger, J., Sasse, M. A., and McCarthy, J. D. The mechanics of trust: A framework for research and design. *International Journal of Human-Computer Studies*, 62(3):381–422, March 2005.

Rotter, J. B. Interpersonal trust, trustworthiness, and gullibility. *American psychologist*, 35(1):1–7, 1980.

Salo, A. Robo advisor, your reliable partner? Building a trustworthy digital investment management service. Master's thesis, 2017.

Schnell, C. Wenn's ums geld geht – männersache. Retrived from https://www.handelsblatt.com/finanzen/maerkte/boerse-inside/privatvermoegen-wenns-ums-geld-geht-maennersache-seite-2/5784664-2.html, 2011.

Scopelliti, M., Giuliani, M. V., and Fornara, F. Robots in a domestic setting: a psychological approach. *Universal access in the information society*, 4(2): 146–155, 2005.

Shandwick, W. Ai-ready or not: Artificial intelligence here we come! - weber shandwick, 2016. URL https://www.webershandwick.com/news/ai-ready-or-not-artificial-intelligence-here-we-come/.

Shapira, Z. *Risk taking: a managerial perspective*. New York, Russel Sage Foundation, 1995.

Siau, K. and Wang, W. Building trust in artificial intelligence, machine learning, and robotics. *Cutter Business Technology Journal*, 31(2):47–53, 2018.

Slovic, P. and Lichtenstein, S. Relative importance of probabilities and payoffs in risk taking. *Journal of Experimental Psychology*, 78:1–18, 1968.

Statista. Robo-Advisors Deutschland. Technical report, 2018.

Statman, M. Behaviorial finance: Past battles and future engagements. *Financial Analysts Journal*, 55(6):18–27, November 1999.

Taddeo, M. Modelling trust in artificial agents, a first step toward the analysis of e-trust. *Minds and machines*, 20(2):243–257, 2010.

Teigen, K. H. Risk-taking behavior. *Journal of Behavioral Decision Making*, 9:73–74, March 1996.

Thelisson, E. Towards Trust, Transparency and Liability in AI/AS systems. In *IJCAI*, pages 5215–5216, 2017.

Thelisson, E., Padh, K., and Celis, L. E. Regulatory mechanisms and algorithms towards trust in AI/ML. In *Proceedings of the IJCAI 2017 Workshop on Explainable Artificial Intelligence (XAI), Melbourne, Australia*, 2017.

UN. World population prospects: The 2012 revision. Technical report, 2013.

Van Thiel, D. and Van Raaij, F. Explaining customer experience of digital financial advice. *Economics*, 5(1):69–84, 2017.

Venkatesh, V. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2):186–204, December 2000.

Yousafzai, S. Y., Pallister, J. G., and Foxall, G. R. A proposed model of e-trust for electronic banking. *Technovation*, 23(11):847–860, 2003.