IN BOTS WE (DIS)TRUST?

Master thesis by Håvard Tonning and Martin Underhaug

Universitetet i Stavanger

UIS BUSINESS SCHOOL

MASTER'S THESIS

STUDY PROGRAM:	THESIS IS WRITTEN IN THE FOLLOWING SPECIALIZATION/SUBJECT:
Master of Science in Business Administration	**Economic Analysis**
	IS THE ASSIGNMENT CONFIDENTIAL? **No** (**NB**! Use the red form for confidential theses)

TITLE:

In bots we (dis)trust?

AUTHOR(S)		SUPERVISOR:
Candidate number:	Name:	Kristoffer Wigestrand Eriksen
4121 …………………	Martin Underhaug …………………………………….	
4112 …………………	Håvard Tonning …………………………………….	

ACKNOWLEDGEMENTS

Over the past six months, we have explored the exciting and challenging fields of behavioral and experimental economics, as well as the interesting, confusing and frustrating concept of trust. The work presented in this thesis could not have been carried out without the help and support of our families and friends. However, some individuals deserve a special mention:

First and foremost, we would like to express our deepest gratitude to our supervisor, Associate Professor Kristoffer Wigestrand Eriksen. We thank him for introducing us to the field of behavioral economics and encouraging us to challenge our self throughout this thesis. The thesis brought numerous challenges during our research, but his door was always open whenever help was needed. We would also like to thank Postdoctoral Fellow Sebastian Fest and Professor Ola Kvaløy for valuable insights and discussions. We are particularly thankful for the technical assistance we received from Sebastian when launching the experiments on Amazon Mechanical Turk's online platform.

Next, we would like to thank UiS Business School for the opportunity to conduct the experiments this thesis builds upon. This work would not have been possible without their financial support.

Some of our friends also deserve our deep sense of appreciation. We thank Christian Kvalvåg and Tormod Høie for their technical support. Moreover, we would like to acknowledge our proofreaders Ervin Hoti and Atul Giri. Christian and Ervin deserves a special mention for their outstanding work.

Finally, we would like to thank the LIONESS Lab community for helping us out when we faced problems due to our limited programming experience. A special thanks to the founders Dr. Marcus Giamattei and Dr. Lucas Molleman for constantly answering our inquiries, even on late Saturday nights.

ABSTRACT

As algorithms have evolved to become alternatives to human decision-makers in several domains, trust in algorithms becomes a crucial research topic. Indeed, research have shown that higher levels of trust lead to more reliance and faster adoption of technological artifacts. The intention of this thesis is to examine if people trust algorithms more than their human counterparts. This is done by constructing two experiments which each explore different manifestations of trust. First, we replicate the well-known trust game by Berg, Dickhaut and McCabe to investigate if people trust unknown individuals more than algorithms ('Study 1', $n = 1{,}600$). Next, we employ the 'Judge-Advisor System'—a paradigm used to study the impact of advice on human judgements—and examine if people rely more on a financial advice emanating from a financial advisor compared to a robo-advisor ('Study 2', $n = 350$). All participants were recruited through the online crowdsourcing platform 'Amazon Mechanical Turk.'

The results from 'Study 1' suggest that people trust algorithms more than people. However, this does not seem to translate to the context of financial advisory ('Study 2'), where the participants relied equally on an advice given by a financial advisor and a robo-advisor. Moreover, age does not seem to affect the level of trust in algorithms nor robo-advisors and trust in algorithms seems to be independent of the information revealed about the algorithm.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Algorithms have been around for centuries. The Babylonians used them to find square roots by hand, Greek mathematicians used them to find an approximation of Pi, the greatest common divisor and prime numbers, and the British used them to decipher German Enigma codes (Chabert, Barbin, Borowczyk, Guillemot & Michel-Pajus, 1999; Das, 2016). Decades later, algorithms of the present form, driven by the proliferation of 'big data' feeding into advanced technology, are increasingly making decisions and giving advice in areas that require human judgement (Science and Technology Committee, 2018).

Examples abound. In the health sector, algorithms are employed to assess the risk of cancer, support complex treatment decisions and ensure earlier and more accurate diagnoses (Science and Technology Committee, 2018). In the criminal justice system, they are used to help judges in parole and sentencing decisions by making predictions on the future risk of re-offending (Kehl, Guo & Kessler, 2017; Science and Technology Committee, 2018). And in the recruitment industry, automatic vetting systems are screening candidates and rejecting up to 75% of résumés before a human sees them (Millar, 2012; The Economist, 2018).

Another industry transformed by intelligent algorithms, empowered by the tremendous advancements in computing power, 'machine learning' and 'artificial intelligence', is the financial industry. 'FinTech' investments have never been higher and the banking sector is likely to see more change in the following ten years than it did in the past two centuries (KPMG, 2019; Treanor, 2014). At the same time, consumer preferences are evolving. Customers of financial services are getting more comfortable with computer-generated support and expect banks to leverage their data to create personalized offerings based on their life stage, financial goals and personal needs (Accenture, 2017). In response, wealth management firms are introducing digital financial advisors, known as 'robo-advisors', that utilize mathematical algorithms to invest client assets by automating client advisory.

As more and more complex algorithms and technology continue to penetrate our everyday environments, the role of trust in the human-technology interaction (e.g. trust in algorithms or robots) becomes a crucial research topic. While previous trust literature has focused on trust between humans (e.g. Mayer, Davis, & Schoorman, 1995; Rousseau, Sitkin, Burt, & Camerer, 1998), more recent studies have investigated the concept of trust between humans and technology. Much of this literature has focused on the antecedents and role of trust in such relationships. However, fewer studies have investigated whether humans trust other people more than technology. This is essential to understand as technology in the form of algorithms and robots are increasingly being used as alternatives to human decision-aids in both our personal and professional lives.

By using a game-theoretic framework (a repeated version of 'the prisoners dilemma'), Wu, Paeng, Linder, Valdesolo and Boerkoel (2016) found that humans tend to trust algorithms to a greater degree than other humans. Yet, research on algorithms as decision-aids suggest that people exhibit 'algorithm aversion', a phenomenon where people rely more on an advice given by a human over an advice given by an algorithm (e.g. Dietvorst, Simmons & Massey, 2015; Promberger & Baron, 2006). However, the research on algorithms as decision-aids is ambiguous as other report that people trust algorithmic advice more than human advice (e.g. Logg, Minson & Moore, 2019; Madhavan & Weigmann, 2007).

This thesis aims to elaborate on previous findings and investigate whether humans trust other people more than algorithms. Furthermore, because of the increased utilization of algorithms as decision-aids in financial services and the conflicting results from previous studies—which may be a result of the nature of the tasks being studied (Lee, 2018)—the thesis also seeks to explore how individuals rely on financial advice from a financial advisor as opposed to a robo-advisor. Consequently, two research questions are defined:

RQ 1. Do people trust other people more than algorithms?

RQ 2. Do people rely more on financial advisors or robo-advisors?

To address these questions, two studies are formed. The first study ('Study 1') is based on the well-known 'trust game' (also referred to as the investment game), introduced by Berg, Dickhaut and McCabe (1995) [BDM]. In its most basic form, the trust game consists of two anonymous agents: a trustor (sender) and a trustee (responder). After given a monetary endowment, the trustor is given an option to send all, some or none of the money to the trustee. Any amount sent grows (normally triples) before reaching the trustee. Next, the trustee decides how much of the received amount to return to the trustor. In our replication, different conditions were created to explore how the level of trust, measured by the amount transferred by the trustor, depends on the characteristics of the trustee (labeled as a human or an algorithm).

The second study (‘Study 2’) adopt the ‘Judge-Advisor System’, a paradigm used to study the impact of advice on human judgements (Sniezek & Buckley, 1995). Similar to Önkal, Goodwin, Thomson, Gönül and Pollock (2009), participants were asked to provide a price forecast for different stocks. Subsequently, they received an (identical) advice from either a financial advisor or a robo-advisor depending on which condition they were assigned to. The subjects were then asked to revise their initial estimate, allowing them to weigh the advice relative to their first estimate. By observing which condition that weighted the advice the most, we could determine if there were any effect of source on advice utilization.

The rest of this thesis is structured as follows. Section 2 begins with a review of the trust literature in order to understand the concept of trust, how trust is measured and why people choose to trust. Then we will present literature on trust in technology, followed by the importance of trust within the financial industry. Section 2 ends with an introduction to the ‘Judge-Advisor System’ and previous research on advice utilization. The subsequent section (section 3) incorporates both ‘Study 1’ and ‘Study 2.’ First an overview of the experimental platform is presented, followed by the experimental design, proceedings, predictions, results and a discussion for each study respectively. Section 4 offers a general discussion and the implications of the two studies, while section 5 concludes. Finally, section 6 looks at the thesis’ limitations and illuminates the possibilities for future research.

2. THEORETICAL BACKGROUND

2.1. Towards a definition of trust

Over the past six decades, researchers across multiple academic disciplines, predominantly psychologists, sociologists and economists, have been studying the concept of trust. Yet, little consensus has been formed. Indeed, more than 300 definitions have been proposed, and over 700 articles focusing on trust as their primary research topic have been published (Schaefer, 2013). Some of these are explanatory or conceptual pieces, while others are empirical or experimental (Lyon, Möllering & Saunders, 2011). Some take the perspective of the trustor, while others recognize that to fully understand trust, one must see it in the light of the qualities and behaviors of the trustee (Lewicki & Brinsfield, 2011). Furthermore, some argue that trust is a behavior (e.g. Coleman, 1990; Fehr, 2009; Elster, 2007), while others define trust as a personal disposition (e.g. Rotter, 1967, 1971) or a state of mind (e.g. a belief or an expectation) (e.g. Mayer et al., 1995; Rousseau et al., 1998) (Lewicki & Brinsfield, 2011).

Advocates of the behavioral-based approach to trust argue that 'trust is best seen as ways of acting' (Reiersen, 2017, p. 436). According to Luhmann (1979), trust is a decision taken by a trustor based on familiarity, expectations and risk. Moreover, Coleman (1990) writes about the decision to place trust and compare it with the decision to place a bet. Elster (2007, p. 344) infer that trust is to '…refrain from taking precautions against an interaction partner' and Fehr (2009, p. 238) defines trust as a behavior where an individual 'trusts if she voluntarily places resources at the disposal of another party (the trustee) without any legal commitment from the latter.' From the behavioral standpoint, beliefs and expectations are reasons for which an agent decides to trust, while trust itself is a matter of choice and actions characterized by the way people behave (Reiersen, 2017).

In contrast, the belief-based approach view trust as a belief about others' trustworthiness (Reiersen, 2017), which in turn can be grounded in expectations about others' ability, benevolence, and integrity (Mayer et al., 1995). Gambetta (1988) defines trust in terms of subjective probabilities, Robinson (1996, p. 576) in terms of 'expectations, assumptions or beliefs' and Rousseau et al. (1998, p. 395) argue that 'trust is a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intensions or behavior of another.' Consequently, actions can be seen as a result of trust. Indeed, Bauer (2015, p. 8) distinguishes 'trust' from 'trusting behavior' and states that 'trust is an expectation and not a decision or a behavior.' Moreover, Rousseau et al. (1998, p. 395) claims that 'trust is not a behavior (e.g. cooperation), or a choice (e.g., taking risk), but an underlying psychological condition that can cause or result from such actions.' Hence, the belief-based approach view trust as an antecedent of trusting behavior, not as a behavior itself.

While the belief-based approach to trust argue that trust can be seen as a belief about others' trustworthiness, Dietz and Den Hartog (2006) note that even though a trustor may consider a trustee to be trustworthy, it does not necessarily mean that the trustor actually trusts the trustee. In their view, trust can take three different forms: a belief, a decision and an action. This leads to a three-stage process of trust. In the first stage, the trustor forms a belief about the trustee's trustworthiness. Next, this belief is manifested through the intention of making oneself vulnerable to potentially harmful actions of the trustee (Dietz & Den Hartog, 2006). However, this is not enough as the decision to trust 'only implies an intention to act' (Dietz & Den Hartog, 2006, p. 559). Consequently, Dietz and Den Hartog (2006, p. 559) argue that the trustor must commit themselves to a 'trust- informed, risk-taking behavior' to demonstrate their trust. Thus, trusting behavior is a consequence of the decision to trust, which again is based on a belief about the trustees' trustworthiness. A similar model of trust is presented by McKnight and Chervany (2001), where trust-related behavior is seen as a result of an individual's trusting intentions and trusting beliefs, as people 'tend to translate their beliefs and intentions into actions' (McKnight & Chervany, 2001, p. 39).

The widespread views of trust have led researches to call it an ambiguous and elusive concept (Bauer, 2015; Lyon et al., 2011; Yamagishi & Yamagishi, 1994). McKnight and Chervany (2001) compare the trust literature with the story of the six blind men who together were to explain an elephant by touching different parts and Shapiro (1987, p. 625) call the state of trust definitions a 'confusing potpourri.' Yet, there seems to be an agreement among most scholars that for trust to arise, both risk and interdependence must be present (Rousseau et al., 1998). Mayer et al. (1995, p. 712) remarks that 'trust is not taking risk per se, but rather it is a willingness to take risk', and Hardin (2002, p. 11) notes that '…acting on trust involves giving discretion to another to affect one´s interest. This move inherently subject to the risk that the other will abuse the power of discretion.' Following, vulnerability and expectations seem to be fundamental elements when defining trust (Rousseau et al., 1998). Indeed, Evans and Krueger (2009, p. 1004) note that 'without personal vulnerability, trust devolves into confidence – a belief without consequence.' Moreover, when analyzing 121 definitions of trust, Walterbusch, Gräuler and Teuteberg (2014) found that 47.9% of the definitions included the word 'expectation', while 'vulnerability' was used in 23.1% of the cases.

Considering the discussion above, we take the view that trusting behavior (an action) is a consequence of the decision to trust (trusting intention) based on a belief about others' trustworthiness (trusting beliefs). Thus, this thesis adopts a widely held definition of Mayer et al. (1995, p. 712):

> *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.*

This definition also holds in a human-technology interaction (e.g. human-algorithm or human-robot interaction), although the beliefs about the dependent object's trustworthiness may be based on other factors, such as the characteristics of the technology itself, as well as its perceived performance, reliance, functionality and helpfulness (McKnight, Carter, Thatcher & Clay, 2011). Indeed, McKnight et al. (2011, p. 7) define trust in a specific technology (or more specifically, trusting beliefs) as 'the beliefs that a specific technology has the attributes necessary to perform as expected in a given situation in which negative consequences are possible.' This will be further discussed in section 2.4.

2.2. Measuring trust

Historically, much of the empirical work on measuring trust have drawn on answers from different survey questions similar to the National Opinion Research Center's General Social Survey (GSS): 'Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?' (Sapienza, Toldra-Simats & Zingales, 2013). Such attitudinal questions have attempted to assess trustors' willingness to accept risk or vulnerability along with the trustors' beliefs about trustees' intentions by analyzing peoples' self-reported responses (Lewicki & Brinsfield, 2011). However, Glaeser, Laibson, Scheinkman and Soutter (2000) argue that attitudinal questions measure trustworthiness, not trust. Moreover, to determine whether someone is trusting, Glaeser et al. (2000) advise one to ask them about specific instances of trusting behavior.

In contrast to attitudinal questions, behavioral scholars—primarily psychologist and behavioral economists—have undertaken laboratory experiments constructed as interactive games grounded in game theory to better measure trust by eliciting trusting behavior (e.g. prisoners dilemma and trust games) (Evans & Krueger, 2009). Common for these experiments is that a trustor is given a choice to trust or not. The decision to trust offers a potential gain, but it also makes the trustor vulnerable to the behavior of the trustee (Evans & Krueger, 2009). On the other hand, no trust yields no loss. Although such experimental games primarily measure trusting behavior, we argue that trusting behavior is an indicator of trust as people 'tend to translate their beliefs and intentions into actions' (McKnight & Chervany. 2001, p. 39). Indeed, Evans & Krueger (2009, p. 1004) argue that experimental games 'provide an external, quantifiable measure of the underlying psychological state of trust' and Naef and Schupp (2009) advocates that common trust experiments offer a valid measure of trust in strangers.

2.2.1. The trust game by Berg, Dickhaut and McCabe (1995)

Since the introduction of the trust game by Berg et al. (1995), the experiment has been frequently replicated and come to be the standard experiment to study trust and trustworthiness in behavioral economics (Evans & Krueger, 2009). The trust game consists of two players (a sender and a responder) that are paired anonymously. At the beginning of the game, each player is given an endowment of S (S=$10 in the original BDM experiment). The sender then decides whether to transfer all, some or none of the endowment to the responder. Any amount $s \in [0, S]$ sent by the sender is multiplied by a factor X (typically, X is 3) so that $X * s$ is passed on to the responder. In turn, the responder decides how much to return, $r \in [0, X * s]$, back to the sender. Consequently, the sender earns the endowment, minus the transferred amount, plus any amount returned by the responder $(S - s + r)$. The responder earns the endowment, plus the multiplied amount sent by the sender minus the returned amount $(S + X * s - r)$. The amount sent by the sender is said to capture the degree of trust, while the amount returned by the responder is used as a measure of trustworthiness. A zero transfer is associated with no trust, while a higher amount sent (higher s) indicates greater trust. Similarly, a zero-return amount suggests that the trustee is not trustworthy, while a higher amount returned (higher r) is associated with greater trustworthiness.

Using backward induction, it is evident that the subgame perfect Nash equilibrium in a one-shot anonymous trust game is for the sender to keep the entire endowment. Given that the responder is self-interested, they will take advantage of the sender's vulnerability and retain the entire amount sent by the sender. Anticipating the responder's decision, the sender should keep the endowment and send no money to the responder in the first place. Consequently, neoclassical economic theory, based on the assumption that individuals are rational and purely self-interested, predicts no trust and reciprocation as self-interest undermines trust and discourage reciprocity (Evans & Krueger, 2009). However, observed behavior is quite different. In the original BDM experiment, 55 out of 60 senders transferred an average of $5.65 of their

$10 initial endowment (56.55%) (Berg et al., 1995).[1] Furthermore, multiple replications of the BDM trust game reveal that there is a substantial willingness for senders to make themselves vulnerable to the trustworthiness of the responder by choosing to send a significant amount of their initial endowment (see Johnson & Mislin, 2011 for a review).

Although the trust game is widely used to study trust in behavioral economics, several scholars note that senders' behavior cannot only be explained by beliefs about others' trustworthiness (Ashraf, Bohnet & Piankov, 2006; Sapienza et al., 2013). Cox (2004) argue that the amount sent may be driven by people's altruistic preferences and that individuals may be inclined to send parts of their endowment out of pure kindness. Indeed, Sapienza et al. (2013, p. 1325) note that transfers 'lower or equal to 25% of the initial endowment can be interpreted as an act of charity more than an act of trust.' Moreover, Bohnet and Zeckhauser (2004) suggest that the amount sent might be affected by people's betrayal aversion. In a series of experiments, they found that people sent less of their endowment to another person than a computer which randomly chose how much to return. Consequently, they argue that (when keeping probabilities of outcomes equal) people will require a premium to engage in a risky lottery where the outcome is determined by another person, as opposed to an identical lottery where nature is in charge of the outcome. Finally, other scholars have remarked that the amount sent can be influenced by other factors, such as individual's risk aversion (Karlan, 2005; Schechter, 2007), inequality aversion (Sapienza et al., 2013) and efficiency preferences (Engelmann & Strobel, 2004).

Contrarily, Brülhart and Usunier (2012) found no support for the assertion that altruism is a statistically significant determinant of 'trust-like' behavior. In addition, other researchers argue that trust games measures trusts separate from risk, and that trusting decisions is distinct from risky decisions (Houser, Schunk & Winter, 2010; McCabe, Houser, Ryan, Smith & Trouard, 2001). Finally, even though some of a sender's behavior in the trust game can be explained by people's preferences, beliefs about others' trustworthiness still seems to play an important role

[1] The original BDM experiment had two conditions ('No history' and 'Social history') which together consisted of 60 senders. The 32 senders in the 'No history' condition sent an average of 51.60% ($5.16) of their $10 endowment, while the 28 senders in the 'Social history' condition sent an average of 53.60% ($5.36) of their $10 endowment (Berg et al., 1995).

(Sapienza et al., 2013). Hence, to further understand why people's observed behavior contradicts the predictions of neoclassical economic theory, an important question to ask is…

2.3. Why do people trust others?

Different explanations have been proposed to explain why people trust others. According to Coleman (1990, p. 99), a rational agent will choose to trust 'if the ratio of the chance of gain to the chance of loss is greater than the ratio of the amount of the potential loss to the amount of the potential gain.' Coleman's statement represents a dominating view within the trust literature, especially within economics, where trust is seen a mean to maximize one's own utility (Banu, 2019). Moreover, it has been widely noted that one of the key determinants of trust is the belief about others' trustworthiness. Indeed, Reiersen (2019, p. 19) argue that 'trust is only sustainable in the presence of widespread trustworthiness', Hardin (2002, p. 30) remark that 'the best device for creating trust is to establish and support trustworthiness' and Thielmann and Hilbig (2015, p. 1523) note that 'trust—especially among strangers—is only defensible if one can expect the trusted person (the so-called trustee) to honor rather than betray the trust.'

If trust is driven by perceived trustworthiness, it is essential to understand what makes someone appear trustworthy. According to Hardin's (2002; 2006) theory of trust as encapsulated self-interest, a trustee might be deemed trustworthy because the trustee could take the trustors interests into account as it might yield the trustee a benefit. This benefit could be directly towards the trustee in the future, or it could be a benefit for the trustee that arise from the well-being of the trustor. Hardin's theory provides a reason for the trustee to act trustworthy, as well as a reason for the trustor to believe that the trustee could be trustworthy. Yet, Hardin (2006) argue that one of the most important reasons for encapsulating the trustee's interests with the interests of the trustor is that there is an ongoing relationship and that the trustee would like the relationship to continue. Consequently, since trust as encapsulated self-interest is most relevant to explain trust in repeated interactions, the concept might have difficulties to explain trust in one-shot interactions with strangers.

One reason for trustors to believe that unknown individuals will behave trustworthy in one-shot interactions is the existence of norm driven trustworthiness. Indeed, Bichierri, Xiao and Muldoon (2011) and Reiersen (2019) found that people believe that others are likely to punish someone who do not reciprocate others trust, concluding that acting trustworthy can be considered a norm. If a trustor believes that being trustworthy is a norm, then trust in strangers can in fact be rational 'insofar as the trusting acts as a signal, whose intended effect is to focus the recipient on a reciprocity norm' (Bichierri et al., 2011, p. 172). However, the belief about others' trustworthiness and the existence of a 'trustworthiness norm' can only act as a foundation for trust and does not state how trust is formed for individual trustees.

Literature on generalized trust, that is, the trust we have in unknown individuals, have examined how trust is formed between people with no relation (Dinesen & Bekkers, 2015). It is mainly split between two views: the experiential and the dispositional explanation. Advocates of the experiential explanation argue that trust is a direct consequence of the perception of others' trustworthiness, formed by past experiences and under continuous development (Dinesen & Bekkers, 2015). In contrast, advocates of the dispositional explanation posit that trust is either a 'downstream consequence of proximate dispositions such as personality traits' or a 'facet of personality in its own right' (Dinesen & Bekkers, 2015, p. 5). In either case, the dispositional explanation view trustfulness as a deeply rooted disposition and note that trust is formed through generic transmission or socialization in early stages of life (Dinesen & Bekkers, 2015). Although the experiential and dispositional explanation differ on how trust is formed and its stability through an individual's life, they are not considered mutually exclusive. The concepts might even interact (Dinesen & Bekkers, 2015). As a result, generalized trust could be a product of previous experiences and deeply rooted dispositions, which both will affect the beliefs about others trustworthiness, and consequently, the trust towards others.

As illuminated above, various theories have been proposed to explain why people trust others. However, as humans increasingly interacts with technological artifacts, it is important to understand that trust in people and trust in technology might differ.

2.4. Trust in technology

Although some researchers have argued that people cannot trust technology (e.g. Friedman, Khan & Howe, 2000; Sceidermann, 2000), an increasing number of scholars recognize the existence of human-technology trust (Lankton, McKnight & Tripp, 2015). Indeed, trust in technology have become a diverse research domain with papers ranging from trust in automation (see Lee & See, 2004 for a review) to trust in robots (see Hancock, Billings, Schaefer, Schen, de Visser & Parasuraman, 2011 for a review) and online recommendation agents (e.g. Komiak & Benbasat, 2006; Wang & Benbasat, 2005). Moreover, previous studies have shown that trust in technology affect people's strategy towards the use of technology (Bagheri & Jamieson, 2004; Muir, 1987). In addition, Lee and See (2004) note that a higher level of trust leads people to rely more on automated artifacts. This has led researchers to integrate trust into the 'Technology Acceptance Model' (TAM), a model that seeks to explain why people choose to accept or reject new technologies (Wu, Zhao, Zhu, Tan & Zheng, 2011; Xu, Le, Deitermann & Montague, 2014).

One should note, however, that many of these studies have examined trust in technology based on human trust attributes, such as ability, benevolence and integrity (McKnight et al., 2011). Although insights from the interpersonal trust literature is important when studying trust in technology (Schaefer, 2013), McKnight et al. (2011) argue that trust in technology and trust in people are two related, yet distinct constructs. According to them, a major difference between trust in people and trust in technology is that people are 'moral and volitional agents', while technology is 'human-created artifacts with a limited range of capabilities that lacks volition and moral agency' (McKnight et al., 2011, p. 4).

In line with the trust definition given by Mayer et al. (1995), McKnight et al. (2011, p. 4) view trust as 'beliefs that a person or a technology has the attributes necessary to perform as expected in a situation.' However, trust in people and trust in technology differ in that people´s beliefs about the dependent object are based on different attributes. Indeed, McKnight et al. (2011, p. 9) suggest that trust in technology is based on beliefs about the technology's functionality, helpfulness and reliability, in which functionality is the technology's 'capacity or capability to complete a required task', helpfulness is a feature of the technology itself and reliability is the technology's capacity to perform consistently and predictably. Despite the distinction, both the human-like approach (ability, benevolence and integrity) and the technology-like approach (functionality, helpfulness and reliability) have been used in previous trust studies depending on the nature of the technology. For example, Lankton et al. (2015) note that it seems reasonable to use the human-like approach when studying more human-like technology such as online recommendation agents. Similarly, the technology-like approach seems reasonable when studying technology with fewer human traits.

Another aspect of trust in technology is the fact that technologies are designed, operated and controlled by humans. Thus, even though technology does not exhibit their own will or moral compass, it is created with an objective and therefore seek to fulfill goals of the designers (Lee & See, 2004). Following, Sztompka (1999) claims that trust in people and trust in technology are essentially the same construct, as people bestow their trust to the creators and operators of the technology (Wang & Benbasat, 2005). Other studies report that humans respond socially to technology by treating technological artifacts similar to people, rather than simple tools (Reeves

& Nass, 1996). Indeed, Wang and Benbasat (2005, p. 72) found that people perceive 'human characteristics (…) in computerized agents' and 'treat online recommendation agents as social actors.' Yet, Lee and See (2004, p. 66) note that trust in people and trust in technology differ as interpersonal trust 'is often a part of a social exchange relationship.' They argue that there is symmetry in trust between people as the involved parties are aware of each other's behavior, intentions and trust. This symmetry is not present in a human-technology relationship, which may affect how people trust technology (Lee & See, 2004). Furthermore, McKnight et al. (2011) stress the importance of distinguishing trust in people and trust in technology to differentiate beliefs towards the designer of the technology and the cognitions about the features of the technology itself.

From the discussion above, trust in technology is a product of the beliefs about designers of the technology, as well as the characteristics and features of the technological artifact itself. For example, the trust people put in a robo-advisor can be affected by beliefs about the advisor's functionality, helpfulness and reliability, as well as beliefs about the specific supplier of the robo-advisor (e.g. a bank or a wealth management firm).

2.5. The importance of trust in financial advisory

In the aftermath of the 2008 financial crisis, the financial industry suffered a great loss of trust from the general population. People started to question the stability of the financial systems, the validity of the underlying principles and the agents present in the financial markets. Indeed, survey results showed that people's trust towards the stock market, banks and financial professionals (e.g. bankers, brokers and financial advisors) plummeted to a lower level than the trust people reported in random, unknown individuals (Guiso, 2010). The following recession illustrated the importance of trust in the financial industry, an industry that acts as custodians of people's savings. Sapienza and Zingales (2012, p. 124) called it a 'trust crisis' and noted that 'while trust is fundamental to all trade and investment, it is particularly important in financial markets, where people part with their money in exchange for promises.' Moreover, Knights, Noble, Vurdubakis and Willmott (2001, p. 318) remark that 'financial services can be said to be in, or even be, the business of trust' and argue that trust is a fundamental condition for the existence of financial services.

While trust is acknowledged as a fundamental element for the overall functioning of financial systems, it is also essential for investors, especially for retail (i.e. non-expert) investors. Indeed, investing in financial assets and utilizing financial advice involves making oneself vulnerable and accept the risk of potential losses based on expectations of positive returns. Furthermore, commissions are often complex, while sales-based incentives may lead to biased advice (Lahance & Tang, 2012). Hence, consumers of financial service providers need to be confident that the financial markets are fair and that the financial institutions and professionals do not exploit the vulnerable position the investors put themselves in. Moreover, non-experts often lack financial literacy and therefore seek financial professionals to get a better insight into the diverse investment options that exist (Van Raaij, 2016). Yet, Sunikka, Peura-Kapanen and Raijas (2010) found that consumers consider financial advisors to be more loyal to their employer than their clients. Thus, trust is essential in a client-advisor relationship. The findings of several researchers support this sentiment: Burke and Hung (2016) note that trust is a key determinant of seeking financial professionals for advice, Guiso, Sapienza and Zingales (2008) find that stock market participation increases with trust and Sunikka et al. (2010) report that consumers have higher trust in their own financial advisor than financial advisors in general, indicating that they selectively choose financial advisors they believe they can trust.

The perceived trustworthiness of a financial advisor can be affected by individual beliefs about the advisor's benevolence, integrity and ability, as well as shared values and effective communication (Ennew & Sekhon, 2007). With large sums of money at stake and significant investment risk present, investors need to believe that the advisor is concerned about their interests and acts accordingly. Furthermore, when utilizing financial advice, people rely on the advisors' expertise and financial knowledge. Indeed, Sunikka et al. (2010) find that trust in competence, integrity and benevolence are the most important characteristics of a financial advisor. Moreover, Madamba and Utkus (2017, p. 5) found that emotional factors 'that bring about positive feelings or sensibilities in the investor' accounted for 53% of the overall trust in financial advisors, ethical factors such as absence of conflict of interest, reasonable fees and acting in the clients' best interest accounted for 30% of the total trust, while functional factors such as the advisors' credentials, expertise and skills accounted for the remaining 17%.

With new technology and robo-advisors entering the financial industry, the nature and role of trust may change. Reichheld and Schefter (2000) argue that trust is even more crucial in digital rather than physical environments and suggest that the need for trust emerges from the lack of human interaction. However, while previous research report that human characteristics is an important driver of trust in financial advisors, a recent study by Hodge, Mendoza and Sinha (2018) found that people's inclination to rely on financial advice provided by a robo-advisor decreases when the robo-advisor is given human attributes. Specifically, naming the robo-advisor reduces the likelihood that investors follow their advice. As previously mentioned, trust in technological attributes and trust in people are two related, yet distinct constructs. When evaluating robo-advisors' trustworthiness, investors might focus more on the advisors' functionality, helpfulness and reliability, as well as the trustworthiness of the developer and financial institution associated with the advisor. Indeed, Yousefi and Naisiripour (2015) found that the features of the bank had the greatest impact on customers' trust in e-banking services. Previous research also suggest that people trust computers to provide more unbiased information than humans (e.g. Fogg, 2009; Fogg & Tseng, 1999). Hence, functional factors might have a greater influence of the overall trust in financial advisors (both human and robo-advisors) as new agents continue to emerge.

2.6. Trust and advice utilization

When making important decisions, people often rely on advice from various sources with the expectation that the advice can reduce their uncertainty and improve their judgement. In doing so, decision-makers make themselves vulnerable to the competence and intentions of the advisor (Van Swol & Sniezek, 2005). Consequently, relying on advisors and utilizing advice is often associated with trust. Indeed, Doney, Cannon and Mullen (1998, p. 604) define trust as a 'willingness to rely on another party and to take action in circumstances where such action makes one vulnerable to the other party.' Moreover, previous research on advice utilization report a strong relationship between trust and the degree to which an advice is taken into account. For example, Sniezek and Van Swol (2001) find that trust increases the likelihood of taking an advice, Jungermann and Fischer (2005) note that people largely rely on their trust in the advisor when deciding to accept or reject advice and Prahl and Van Swol (2017) argue that advice utilization is a behavioral measure of trust.

2.6.1. The Judge-Advisor System

To examine how people utilize advice, researchers on judgement and decision-making have often employed the 'Judge-Advisor System' (hereafter, 'JAS'). A 'typical' JAS study consists of a judge (the decision-maker) and an advisor. First, the judge is asked to provide an initial decision before being presented with a recommendation from an advisor. Next, the judge must decide to follow the advice or not. Importantly, they are under no obligation to follow the advisor's recommendation and can therefore choose whether to take the recommendation into consideration or not (Bonaccio & Dalal, 2006). In some studies, the advice is dichotomous (accept or reject), while in others, the judge can adjust their initial decision towards the advisor's recommendation (e.g. forecasting tasks) (Bonaccio & Dalal, 2006). Adjusting the final decision towards the advice is referred to as advice utilization, while advice discounting exist if a judge chooses not to follow advice, but rather follow their own instincts (Bonaccio & Dalal, 2006).

Several findings are worth noting from the JAS literature (see Bonnacio & Dalal, 2006 for a complete review). Despite the fact that following advice generally helps judges make better decisions, multiple studies have found evidence of 'egocentric advice discounting', a phenomenon where people 'overweigh their own opinion relative to that of their advisor' (Bonnacio & Dalal, 2006, p. 129). Harvey and Fischer (1997) claim that egocentric advice discounting occurs because people are overconfident in their own abilities and anchored towards their initial estimates, while Yaniv and Kleinberger (2000) note that people have full access to their own thoughts and reasonings' and less information about the advisor's. Another finding is that advice utilization increases with the advisors' perceived expertise (Bonnacio & Dalal, 2006; Jungermann & Fischer, 2005). Related is the finding that people are more likely to follow what they perceive to be a good advice compared to what they see as a poor advice (Yaniv & Kleinberger, 2000). Moreover, Gino and Moore (2006) find that advice utilization increases with the complexity of the task (see also Schrah, Dalal & Sniezek, 2006), while Bonaccio and Dalal (2006) note that it decreases if the judge questions the intentions of the advisor. Hence, it is argued that trust in the advisor is an important determinant of advice utilization (Sniezek & Van Swol, 2001; Van Swol & Sniezek, 2005; Prahl & Van Swol, 2017). Finally, Heath and Gonzalez (1995) find that receiving an advice increases judges' confidence in their final decisions and Van Swol (2009) report that judges' confidence is strongly correlated with how much they trust an advice.

The majority of these studies have investigated how people react to the advice from human sources. However, as algorithms, computers and expert systems have evolved to become alternatives to human advisors, researchers have started to examine the degree to which people rely on advice from non-human sources.

2.6.2. Relying on advice from non-human sources

Researchers studying advice utilization have also investigated how people rely on advice that emanates from non-human advisors. Several domains have been investigated, ranging from medical recommendations (e.g. Promberger & Baron, 2006) and financial recommendations (e.g. Önkal, 2009), to more subjective domains like humor and attractiveness (e.g Yeomans, Shah, Mullainathan & Kleinberg, 2019; Logg et al., 2019). Results are ambiguous and seem to depend on the task under investigation (Lee, 2018). In the medical domain, people seem to prefer an advice from a medical professional as opposed to a computer program, even though the computer program is more likely to provide a better advice (Promberger & Baron, 2006). Yeomans et al. (2019) also found algorithm depreciation when studying joke recommendations. People relied more on advice from friends rather than algorithms. In contrast, Logg et al. (2019) found that people utilize advice more when it comes from an algorithm than when it comes from a person. They studied advice utilization through several domains, including estimation of people's weight, popularity of songs and attractiveness. In all their experiments, Logg et al. (2019) found evidence of advice appreciation.

In the domain of financial forecasting, Önkal et al. (2009) studied how subjects utilized advice from human experts versus statistical methods when presented with a financial forecasting task. The findings indicate that people rely more on the advice given by the human experts. For forecasting tasks in other domains, Dietvorst et al. (2015) investigated how people utilized advice after seeing the algorithm perform. The findings suggest that after seeing an algorithmic advisor err, the algorithm is punished harder than a human advisor. Consequently, people seem to tolerate mistakes from human advisors more than algorithmic advisors. In fact, the results showed that after observing an algorithmic advisor outperform a human advisor, people were still more willing to depend on the human advisor (Dietvorst et al., 2015).

One suggested explanation of the tendency to rely more on human advisors, despite the fact that non-human advisors like statistical methods, computer programs and algorithms are often more precise than human expertise (e.g. Meehl, 1954; Dawes, 1979), is that human advisors can be accountable for their recommendations. Relying on a human's advice therefore shifts the responsibility of the decision, as human advisors can be blamed for their inaccurate precision (Harvey & Fischer, 1997).

3. EXPERIMENTAL PROCEDURE AND RESULTS

3.1. Recruitment, Experimental Platform and Challenges

To answer the research questions given in section 1,[2] we designed two online experiments by utilizing the web-based software 'LIONESS Lab.'[3] Participants were recruited through the online crowdsourcing platform 'Amazon Mechanical Turk' (hereafter 'MTurk') as it offers an active and diverse subject pool with over 500,000 workers at a lower cost than traditional laboratory experiments conducted at university campuses (Arechar, Gächter & Molleman, 2018; Berinsky, Huber & Lenz, 2012; Mason & Suri, 2012). Although participants recruited on MTurk are paid considerably less than subjects in lab experiments, research shows that as long as stakes are present, the size of the stakes does not seem to have a significant impact on subjects' behavior—with the exception of extremely large stakes (Amir, Rand & Gal, 2012; Raihani, Mace & Lamba, 2013). Furthermore, online replications of classical psychology and economics experiments have shown results comparable to those obtained in offline environments (Arechar et al., 2018; Mason & Suri, 2012). As such, the data obtained online seems to be of the same quality as that obtained through traditional laboratory experiments. In fact, Mason and Suri (2012, p. 4) note that 'while there are clearly differences between Mechanical Turk and offline contexts, evidence that Mechanical Turk is a valid means of collecting data is consistent and continues to accumulate' and Berinsky et al. (2012) conclude that 'the MTurk subject pool is no worse than convenience samples used by other researchers in political science.' Yet, despite its benefits, online experiments have some challenges that are normally not present in offline environments (Arechar et al., 2018).

A major challenge for online experiments is participant dropout rates (Arechar et al., 2018). In contrast to physical laboratory experiments, where subjects typically stay till the end, online participants are considerably more likely to abort in the middle of a session. Consequently, online experiments (particularly those with live interaction) are exposed to higher dropout rates

---

[2] RQ1: Do people trust other people more than algorithms?
RQ2: Do people rely more on financial advisors or robo-advisors?

[3] LIONESS (Live Interaction Online Experimental Server Software) is a free software that offers a basic architecture to conduct online experiments with live interaction. It is a shared project between the University of Nottingham, the University of Passau and the Max Planck Institute for Human Development in Berlin (https://www.lioness-lab.org/)

(Arechar et al., 2018). This raises two concerns, the latter more serious than the former. First, dropouts due to exogenous circumstances such as technical issues or random distractions can raise the cost of the experiment. Second, selective dropouts or dropouts contingent on the conditions of the experiment could compromise the collected data which may raise a concern about the internal validity of the experiment (Arechar et al., 2018; Zhou & Fishbach, 2016). This issue was addressed by implementing various procedures to reduce the dropout rates in each experiment. These procedures will be explained in further details below.

Another concern raised about online experiments is that it reduces the experimenters' control of the experimental session. Some studies require that subjects only participate once as the presence of re-takers could violate the assumption of independent observations, which may jeopardize the quality of the collected data (Arechar et al., 2018; Jilke, Van Ryzin & Van de Walle, 2016). While this can be easily controlled in offline environments, detecting re-takers in online sessions requires additional measures (Arechar et al., 2018). To prevent duplicate participation, the employed software (LIONESS Lab) logged the participants IP-address and blocked individuals that had already entered the experimental pages (Lioness Lab, 2018).[4] Moreover, each participants' MTurk worker ID was recorded to prevent workers who had already completed the experiment to join a later session and to ensure that each submitted HIT (Human Intelligence Tasks) had a unique worker-ID.[5]

When participating in experiments at university campuses, participants may ask the experimenters questions if anything is unclear. However, this is difficult to implement in online sessions (Arechar et al., 2018). Consequently, to verify if the subjects understood the task they were to perform, they had to complete a set of comprehension question before continuing to the decision-making phase of the experiments. However, to reduce dropouts and avoid selection bias in our sample, participants in both studies were allowed to take part of the experiment even

[4] To protect the participants' personal data, Lioness Lab record IP-addresses in an anonymized way. The participants' actual IP-address cannot be retrieved by the experimenters (Lioness Lab, 2018).

[5] The Amazon MTurk worker ID is a 14-character alphanumeric code that does not offer any information about a worker's identity (UC Berkeley Committee for Protection of Human Subjects [CPHS], 2018).

if they did not answer the comprehension questions correctly after three attempts.[6] To control for this ‘lack of understanding’, participants who failed to answer correctly where registered and assigned a dummy variable to use in our analysis.

Other measures were also taken to secure the quality of the experiment. A reCAPTCHA was implemented to avoid the use of bots.[7] The reCAPTCHA blocked 63 requests (59 in ‘Study 1’ and 4 in ‘Study 2’), preventing denied subjects to participate in the experiment. Furthermore, specific requirements were set to secure the quality of the participants. Based on previous research and recommendations from other researchers, we required participants to have a minimum of 1,000 approved MTurk HITs and an approval rate of minimum 98% (Amazon MTurk, 2019; Kaufmann & Tummers, 2017). In addition, the geographical location of the participants was restricted to the United States and the minimum age was set to 18. To further reduce the likelihood of dropouts, the HIT descriptions on MTurk gave an approximation of how long it would take to complete the experiment, as well as information about privacy concerns and informed consent. This reduced the time the participants spent in the experimental pages. Importantly, the HIT description revealed no detailed information about the experiment or the decision task to reduce selection bias.

3.2. Study 1: Trust Game—Berg et al. (1995)

3.2.1. Experimental design

To address the first research question ‘Do people trust other people more than algorithms?’ we replicated the original BDM trust game and manipulated the characteristics of the trustee (a human or an algorithm) by constructing four conditions (one baseline and three treatments):[8]

Strategy Method (‘SM’) – The baseline group consisted of a sender playing the trust game against a responder (another MTurk worker) who had already provided a conditional response for every possible amount sent by the sender (i.e. the responder employed the strategy method).

Game Method (‘GM’) – The first treatment group differed from the baseline (SM) in that the sender and responder (also another MTurk worker) were playing the trust game simultaneously,

6 This was not revealed to the participants.

7 reCAPTCHA is a service offered by Google to identify bots and protect websites from spam and abuse (Google, n.d.).

8 The instructions and the hyperlinks for each condition are attached in Appendix A.

providing their decisions sequentially, similar to the original BDM experiment (i.e. the responder employed the game method, also referred to as the direct-response method).

Transparent Algorithm ('TA') – The second treatment group was identical to the baseline (SM) except that the senders were informed that the responder was a pre-programmed algorithm programmed by another MTurk worker.

Black-Box Algorithm ('BBA') – The third treatment was similar to the transparent algorithm treatment (TA) except that the senders were not given any information about how the algorithm was programmed or who programmed it.

Henceforth, the strategy method condition (SM) and the game method treatment (GM) will jointly be referred to as the human conditions, while the transparent algorithm treatment (TA) and the black-box algorithm treatment (BBA) will be addressed as the algorithmic treatments.

By manipulating the characteristics of the trustee (a human or an algorithm), we investigated if people trust other people more than algorithms. This was done by comparing the amount sent by senders in the strategy method condition versus the transparent algorithm treatment. These conditions were essentially identical in that the responder in both conditions employed the strategy method (made conditional responses for every possible amount sent by the sender). However, senders in the strategy method condition were told that they were playing against another person, while the senders in the transparent algorithm treatment were told they were playing against an algorithm pre-programmed by another person.[9] Hence, the only difference between the conditions were the framing of the responder (see instructions in Appendix A). The black-box algorithm treatment was added to examine if the information revealed about the algorithm affected participants' behavior, while the game method treatment was included to control for the different elicitation methods in the human conditions (strategy method versus

---

[9] Responders in the strategy method condition were asked to give their conditional responses to every possible amount sent by the sender, while responders in the transparent algorithm treatment (and the black-box algorithm treatment) were asked to provide their conditional responses which would be used to pre-program an algorithm (for further details, see Appendix A). Thus, the participants were not deceived.

game method).[10] Moreover, both the black-box algorithm and the game method treatment was used to check the robustness of the possible findings.

In our version of the trust game, subjects were endowed with $1.00 which was framed in cents (100 cents rather than $1.00) to increase the perceived stakes. For practical reasons (as most responders employed the strategy method), participants could only send and return amounts in increments of five cents. Similar to the BDM experiment, the amount sent was tripled before it reached the responder (Berg et al., 1995). Finally, although the trust game allows researchers to study the behavior of both senders and responders, this thesis aims to investigate differences in trust across conditions. Consequently, decisions made by responders were excluded from the analysis.

3.2.2. Proceedings

'Study 1' consisted of two experiments: a pre-experiment and a main experiment. The pre-experiment was completed by 600 participants and conducted between 26th and 28th of March to collect the conditional decisions of the responders in the strategy method condition, as well as the algorithmic treatments. These participants played the trust game using the strategy method and were given different instructions depending on which condition they were assigned to (see Appendix A).

As 600 responders had already provided their conditional response, and therefore had to be paired with a sender, the number of participants for each condition in the main experiment were set in advance. Hence, the main experiment (hereafter 'the experiment') was completed by 1,000 participants (800 senders and 200 responders). The experiment was conducted between 1st and 4th of April. Since previous research have shown that MTurk recruitments decreases with time, we divided the experiment into four experimental sessions (not including a few follow-up experiments to match all the pre-collected responders with a sender). Moreover, as one condition relied on live interaction (the game method treatment), it was important to

[10] We recognize that there is more than one factor that differentiate some of the conditions (SM vs. BBA, GM vs. TA and GM vs. BBA). However, only one factor differentiates the SM and the TA (responder framed as human or algorithm),) the TA and the BBA (information revealed about the algorithm) and the SM and the GM (elicitation method) (see Appendix A for instructions for each condition). Introducing live-interaction for the game method treatment (GM) also implied a different experimental flow than the other conditions, as illustrated in figure 3-1.

maintain a high and steady stream of participants entering the experiment to reduce dropout rates. Thus, participants were given 20 minutes to join a session and 40 minutes to complete the experiment and submit a survey code. Furthermore, we encouraged the participants to start the experiment immediately after accepting the HIT. Each session had high recruitment rates, which reduced the waiting time for subjects in the game method treatment and secured an efficient matching process.

Before accepting to participate in the experiment, the subjects were told that the experiment would take approximately 15 minutes to complete. In addition, they were informed that they would get a $1.00 show-up fee to complete the task, as well as a possible bonus depending on the outcome of the experiment. The size of the bonus was not revealed. Moreover, subjects were informed that by proceeding to the experimental pages, they would give informed consent that their answers could be used for research purposes only.[11] After accepting the HIT, subjects were forwarded to a reCAPTCHA which had to be passed to enter the experiment. At the beginning of the experiment, participants were randomly assigned to either the role of a sender in one of the four conditions, or as a responder in the game method treatment.

Once the participants entered the welcome page, they received some general information about the experimental session. In the next stage, all participants were given the same description and details about the trust game to ensure common knowledge. In addition, they were informed that a summary of the game would be available at every stage of the experiment. In the third stage, participants received specific instructions about the task they were to perform, as well as some information detailing who they were playing against.[12] After reading the instructions, participants were forwarded to a set of comprehension question which had to be answered before entering the decision phase (after three attempts, participants were automatically forwarded to the next stage).

Subjects that were not dependent on live interaction were then forwarded to the decision phase, while those who were assigned to partake in live interaction (game method treatment) were directed to a lobby where they had to wait to be matched with another participant. Once matched, they were sent to the decision phase of the experiment. All participants were given

[11] The Amazon MTurk HIT description is attached in Appendix A.

[12] The instructions for each stage are attached in Appendix A.

two minutes to make a decision. This was important to reduce the waiting time for participants in the game method treatment, which could further reduce the likelihood of participant dropouts. Senders were told that they could send an amount between 0 and 100 cents in increments of five cent. Responders in the GM could return an amount between zero cents and triple the amount sent by the sender in increments of five cents. After making a decision, participants were forwarded to a post-experiment questionnaire. Upon completion, participants were able to see the results and given a survey code to submit on MTurk in order to receive payment.

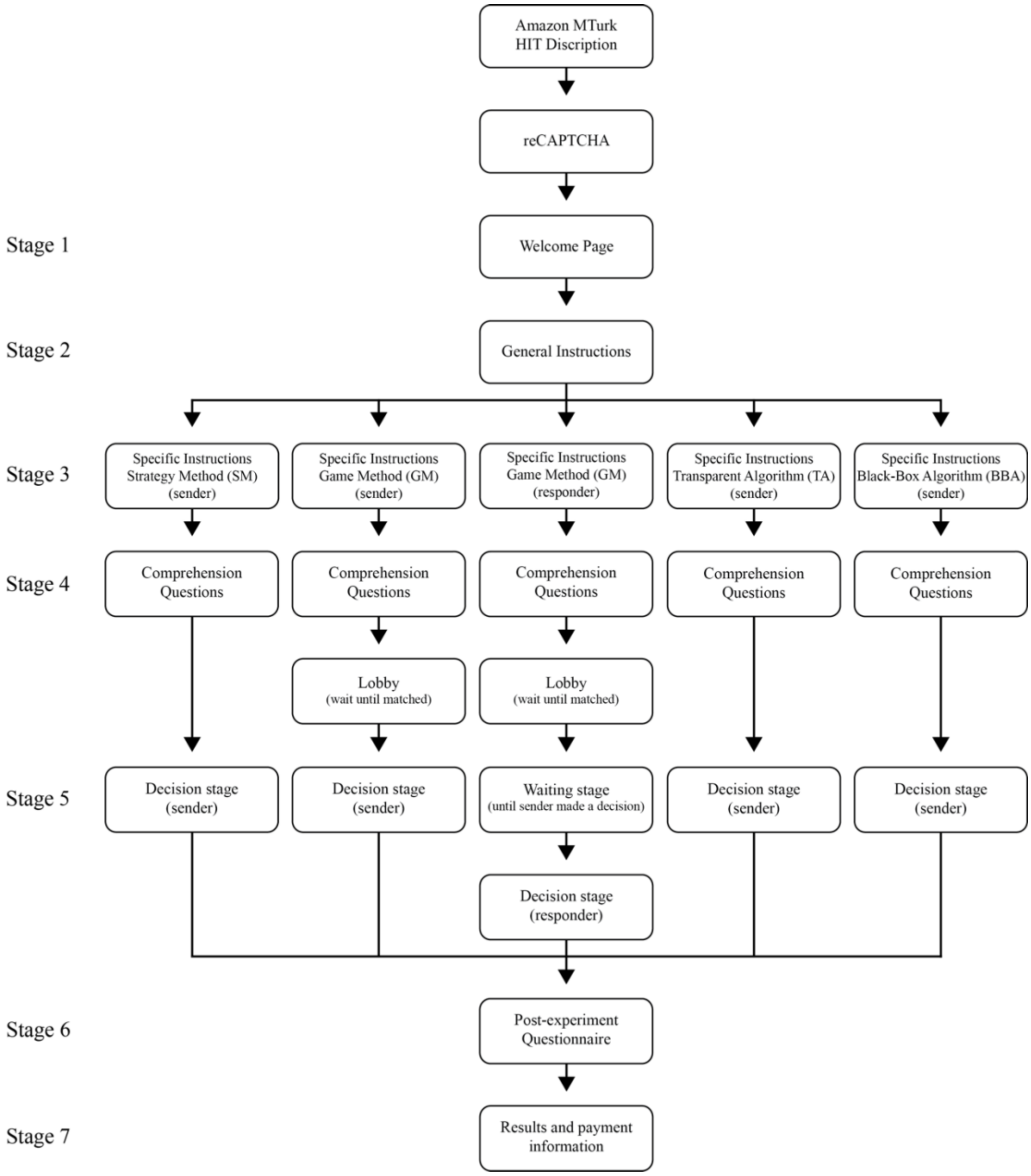


Figure 3-1: Flowchart 'Study 1'

Figure 3-1 illustrates the experimental flow for the main experiment in 'Study 1.' Participants were randomly assigned to the role of a sender in one of the four conditions, or as a responder in the game method treatment.

3.2.3. Dropouts

Despite our measures to prevent participation dropouts, 12.89% (118 of 918) of the senders that entered the experimental pages left before completing the experiment. A closer examination of the dropout data reveals that most of the these left the experiment before the decision phase (see figure 3-2). 30 senders did not proceed past the introductory pages (stage 1–3),[13] and 68 out of the 888 senders that reached the comprehension questions left without submitting a correct answer. Furthermore, seven senders could not be matched with a responder and were therefore unable to proceed to the decision phase of the experiment. These participants were still paid the $1.00 show-up fee as well as the $1.00 endowment from the trust game. Thus, only 1.6% (13 of 813) of the senders that started the decision phase dropped out at some point in the experiment.

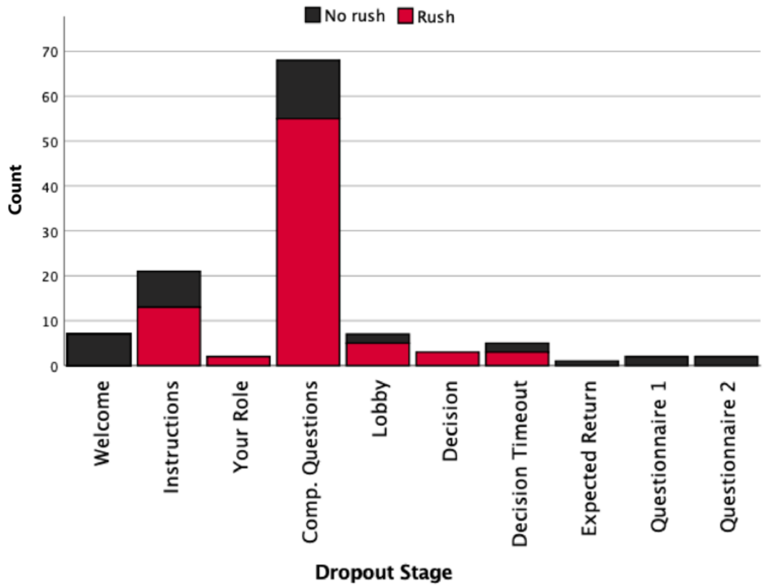


Figure 3-2: Participant dropout by stage

Figure 3-2 shows how many participants that dropped out at each stage of 'Study 1'. The red shaded area illustrates how many participants that rushed through the stage (used 15 seconds or less to complete the stage).

When analyzing the dropout data, we found no indications of selective dropouts. Dropout rates were not statistically different across conditions ($p = 0.582$, Chi-square test of homogeneity, see figure 3-3). Although 10.68% of the senders left the experiment before completing the

[13] Stage 1: Welcome page, Stage 2: General Instructions and Stage 3: Specific Instructions.

comprehension questions, a further analysis revealed that most of these (70 out of 98 or 71.43%) rushed through the introductory pages (stage 1–3), indicating that they did not pay attention to the instructions and were therefore unable to answer the questions successfully (see figure 3-2).[14] Hence, selective dropouts (dropouts depending on the condition of the experiment) did not seem to compromise the quality of the collected data.

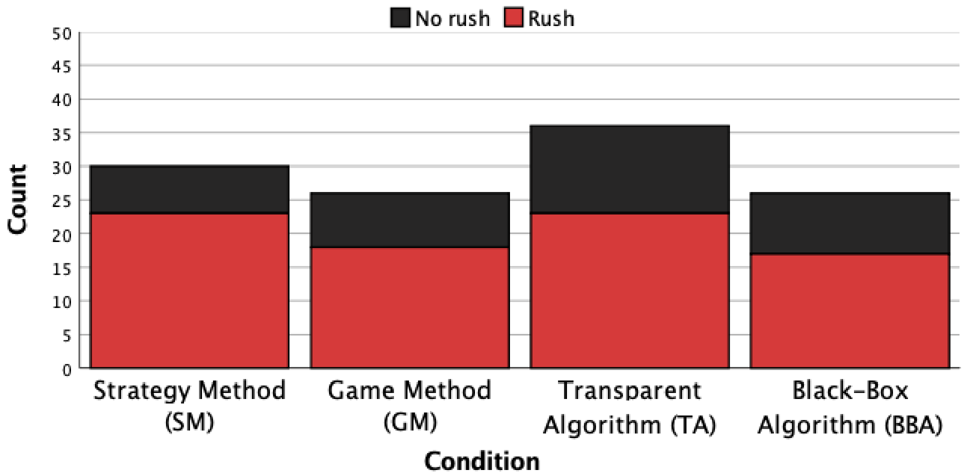


Figure 3-3: Participant dropout by condition

Figure 3-3 shows how many participants that dropped out in each condition. The red shaded area illustrates how many participants that rushed at least one stage during the experiment (used 15 seconds or less to complete at least one stage).

3.2.4. Sample

The final sample consisted of 800 participants (senders) spread across four conditions (200 participants in each condition). The average age of the participants was 37.91 years old ($SD = 11.58$) with the oldest and youngest being 78 and 19 years old respectively. 441 of the participants (55.20%) were male, while 357 participants (44.68%) were female (one participant reported a gender of 'other' and one value is missing).[15] Participants with a bachelor's degree dominated the distribution (40.50%), followed by participants with a high school diploma (20.75%) and technical training (13.75%).[16] 68 participants (8.50%) had a master's degree, while 16 (2%) had completed their PhD. 42.63% of the participants were familiar with the trust

[14] 'Rushed' means that the participant used 15 seconds or less to complete at least one of the introductory pages.

[15] The missing observation is due to a technical error.

[16] Due to a technical error, the education level was not recorded correctly in the first session of the main experiment. Consequently, there are 115 missing observations for the education variable.

game before entering the experiment, and 5.25% of the participants that completed the experiment failed to answer at least one of the comprehension questions correctly.

Table 3-1: Age and gender distribution 'Study 1'

Table 3-1 presents the distribution of age and gender for each condition, as well as for the whole sample. Each condition consists of 200 participants (senders), which yields a total sample of 800 participants.

		Total Sample	Strategy Method	Game Method	Transparent Algorithm	Black-Box Algorithm
Age	Mean	37.91	38.22	37.78	38.06	37.71
	Std. Dev.	11.58	10.84	12.03	11.25	12.21
	Min	19	22	19	21	20
	Max	78	71	70	73	78
Gender	Male	55.20 % (441)	53.27 % (106)	52.50 % (105)	58 % (116)	57 % (114)
	Female	44.68 % (357)	46.73 % (93)	47.50 % (95)	42 % (84)	42.50 % (85)
	Other	0.16 % (1)	0 % (0)	0 % (0)	0 % (0)	0.50 % (1)

3.2.5. Predictions

From a purely rational and self-interested perspective, the dominant strategy in the BDM trust game suggests that people should not trust, as it is irrational for trustees to reciprocate. Yet, Berg et al. (1995) and numerous replications have found evidence of widespread trust and reciprocity, rejecting the hypothesis of neoclassical economists. However, another question arises when comparing trust across groups. Instead of being a question of if people trust, it becomes a whom people trust the most. As the nature of the trust game implies a clear monetary incentive for the trustee to not act in a trustworthy manner, we hypothesize that people will trust another person more than an algorithm. This is based on the finding that people often view algorithms as more rational and objective agents than humans (Dijkstra, Lierbrand & Timminga, 1998), combined with the fact that algorithms to a lesser extent are characterized by the social attributes that makes humans appear trustworthy. Indeed, Lee (2018, p. 3) note that people 'perceive algorithmic decision-makers as more rational, and less intentional and emotional than people.' Hence, our first hypothesis is:

Hypothesis 1.1: People trust other people more than algorithms.

This hypothesis is independent of the employed elicitation method or the information revealed about the algorithm. Consequently, we expect the amount sent by the senders in the strategy method condition and game method treatment to be statistically higher than the amount sent in both the transparent algorithm treatment and the black-box algorithm treatment.

Our second hypothesis address the information revealed about the algorithm. While the trustors in the transparent algorithm treatment were told that the algorithm was pre-programmed by another Amazon Mechanical Turk worker, the trustors in the black-box algorithm treatment were only informed that they were playing against a pre-programmed algorithm (see instructions Appendix A). Consequently, subjects in the black-box algorithm treatment had to rely more on their own beliefs and interpretations of algorithms. Based on the same rationale as for 'Hypothesis 1.1', we predict that the transparency of the algorithm will have a positive effect on trust. Thus, our second hypothesis is:

Hypothesis 1.2: Trust in an algorithm increases with the information revealed about the algorithm.

This implies that the amount sent by trustors in the transparent algorithm treatment is statistically higher than the amount sent by trustors in the black-box algorithm treatment.

Our third hypothesis is based on the observation from the literature on generalized trust which suggests that socialization in early stages of life could affect the propensity to trust other humans. As today's youth are increasingly being exposed to technology from an early age, we predict that younger individuals trust algorithms more than older individuals:

Hypothesis 1.3: Trust in algorithms decrease with age.

3.2.6. Results

'Study 1' sought to investigate how trust differs depending on the characteristics of a trustee. Trust was measured by the amount of money the senders transferred to their responder. Higher amounts sent were associated with a higher level of trust.

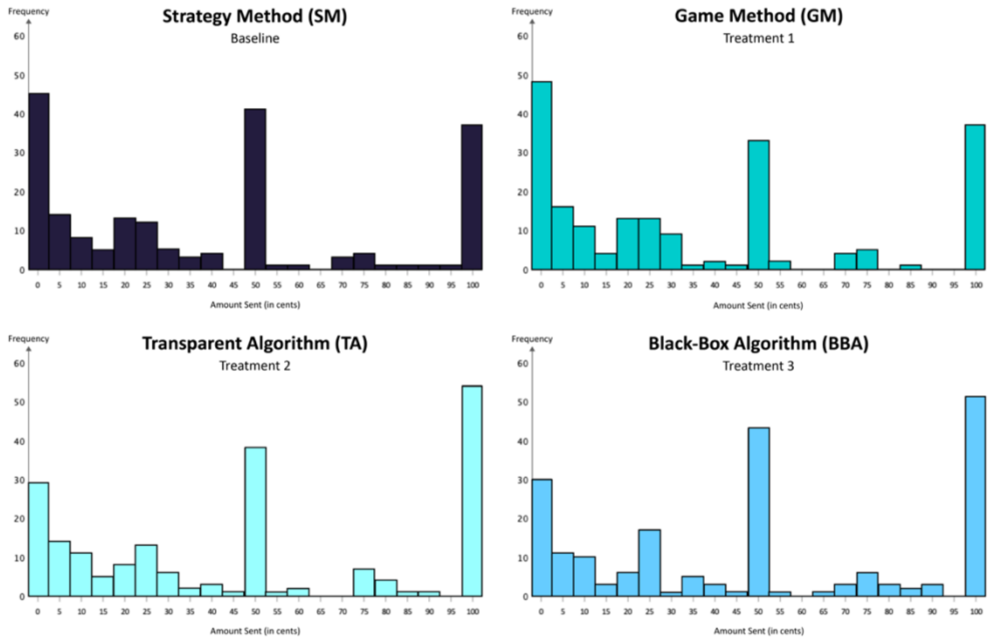


Figure 3-4: Distribution of amount sent

Figure 3-4 illustrates the distribution of the amount sent in cents by the senders in each condition ($n = 200$ in all conditions). The participants could send an amount between 0 and 100 cents in increments of five cents.

Figure 3-4 illustrates the distribution of amount sent across the four conditions. None of the treatments had a significant different distribution compared to the strategy method condition ($p = 0.864$; $p = 0.178$; $p = 0.088$, Kolmogorov-Smirnov Test, see Appendix B). However, by comparing the histograms in figure 3-4 it is apparent that a greater fraction chose to send more in both of the algorithmic treatments, while more subjects chose to send zero in the human conditions. Indeed, figure 3-5 depicts that trustors in the SM and the GM sent on average 39.63% ($SD = 36.33$) and 37.33% ($SD = 36.50$) of their endowment, while trustors in the TA and the BBA sent 48.25% ($SD = 38.07$) and 49.13% ($SD = 37.33$) respectively (see also table 3-2).

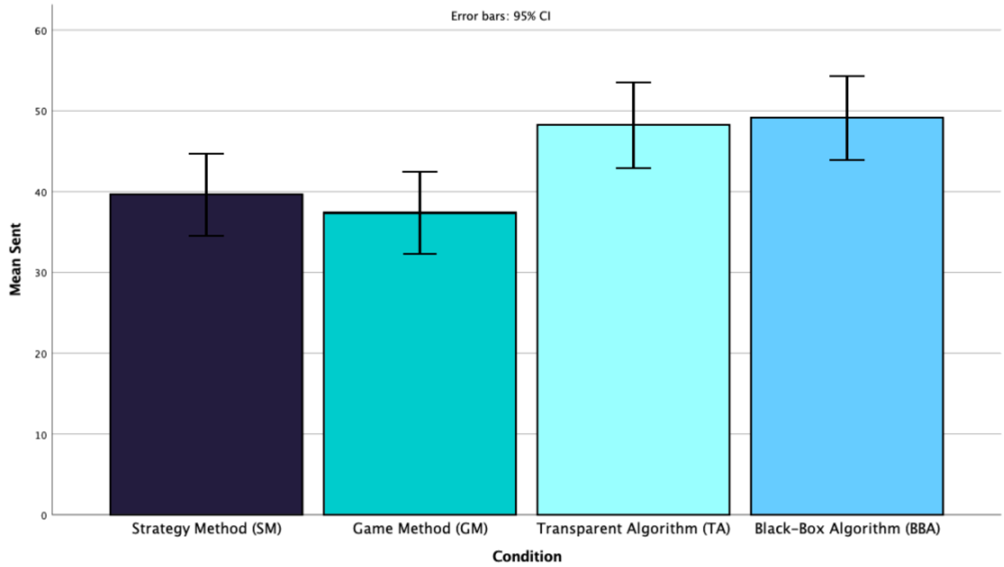


Figure 3-5: Average amount sent by condition

Figure 3-5 illustrates the mean and the 95% confidence interval of the amount sent by the senders in each condition ($n = 200$ in all conditions). The subjects had the option to send an amount between 0 and 100 cents in increments of five cents.

Table 3-2: Descriptive statistics by condition

Table 3-2 presents the mean and standard deviation of the amount sent by senders in each condition ($n = 200$ in all conditions). The subjects had the option to send an amount between 0 and 100 cents in increments of five cents.

	Strategy Method		Game Method		Transparent Algorithm		Black-Box Algorithm	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Amount sent	39.63	36.33	37.33	36.50	48.25	38.07	49.13	37.33

As the distribution of amount sent was not normally distributed in any condition (see the 'Test of Normality: Amount sent', Appendix B), a Mann-Whitney U test was conducted to determine whether the observed differences in the amount sent across conditions were statistically significant.[17] As reported in table 3-3, the amount sent in the TA were statistically significantly higher than the amount sent in the SM, $U = 17\,299.50,\ z = -2.368,\ p = 0.018$. The same holds when comparing the BBA and the SM, $U = 17\,035.50,\ z = -2.600,\ p = 0.009$. Likewise, both trustors in the TA and the BBA transferred a higher amount to their trustees compared to trustors in the GM, $U = 16\,594.50,\ z = -2.984,\ p = 0.003$; $U = 16\,336.00,\ z = -3.211,\ p = 0.001$. Moreover, no significant difference was found between the amount

[17] The Mann Whitney U test is a rank-based nonparametric test that does not require the assumption of normal distributions. It is often presented as the nonparametric alternative to the independent-samples t-test and can be used to determine if there are differences between two groups (Laerd Statistics, 2015).

sent in the SM and the GM, $U = 19\ 258.00,\ z = -0.650,\ p = 0.515$, or the amount sent in the TA and the BBA, $U = 19\ 780.00,\ z = -0.193.,\ p = 0.847$.

Table 3-3: Mann-Whitney U tests amount sent

Table 3-3 presents the output of the Mann-Whitney U tests that were conducted to test differences in the amount sent by the senders in the four conditions ($n = 200$ in all conditions).

Condition	Mean Rank	Sum of Ranks	Mann-Whitney U	z-value	p-value
Strategy Method	187.00	37 399.50	17 299.50	- 2.368	0.018*
Transparent Algorithm	214.00	42 800.50			
Strategy Method	185.68	37 135.50	17 035.50	- 2.600	0.009**
Black-Box Algorithm	215.32	43 064.50			
Game Method	183.47	36 694.50	16 594.50	- 2.984	0.003**
Transparent Algorithm	217.53	43 505.50			
Game Method	182.18	36 436.00	16 336.00	- 3.211	0.001**
Black-Box Algorithm	218.82	43 764.00			
Strategy Method	204.21	40 842.00	19 258.00	- 0.650	0.515
Game Method	196.79	39 358.00			
Transparent Algorithm	199.40	39 880.00	19 780.00	- 0.193	0.847
Black-Box Algorithm	201.60	40 320.00			

* p < 0.05. ** p < 0.01, *** p < 0.001

To isolate the treatment effects (the characteristics of the trustee) from other factors, we conducted two Tobit regressions (upper limit 100 cents).[18] The results of these regressions are presented in table 3-4. Controlling for age and gender did not change the results above. The amount sent by the trustors in the algorithmic conditions were still statistically significantly higher than the amount sent by trustors in the human conditions (model 1). The coefficient for age suggests that the amount sent was not affected by the participants' age ($p = 0.907$), while the gender coefficient imply that females sent more than males ($p = 0.030$). Adding variables for those who already were familiar with the trust game, those who failed the comprehension

[18] The Tobit model is also known as the censored regression model and is used to estimate linear relationships between variables when the dependent variable has either a left- or right-censoring (UCLA Institute for Digital Research & Education, n.d.). In our case, the amount sent was censored at 100 cents (right-censoring).

questions and people's stated risk preferences led to a marginal decrease in the magnitude of the treatment effects (model 2).[19] However, the amount sent in the algorithmic treatments were still significantly higher compared to the human conditions. Moreover, both previous knowledge of the trust game $(p = 0.031)$ and higher risk preferences $(p = 0.009)$ had a positive effect on the amount sent. None of the other variables in model 2 were statistically significant, including the coefficient for females (all $ps > 0.118$).

[19] To control for people's stated risk preferences, we asked the participants the following question: Are you generally a person who takes risk or do you try to evade risks? Please rate your choice ranging between 0 and 10, where 0 is 'Not at all prepared to take risk' and 0 is 'Very much prepared to take risk.'

Table 3-4: Tobit regression on amount sent

Table 3-4 shows the outputs of two Tobit regressions for the amount sent. 'Model 1' is a function of each condition, gender and age, while 'Model 2' also controls for the participants' stated risk preferences, those who were familiar with the trust game before entering the experiment and those who failed the comprehension questions after three attempts. Robust standard errors are reported in the parentheses below the estimated coefficients.

Dependent variable: Amount Sent		
	Tobit Regressions	
Regressor:	(1)	(2)
Condition (ref: Strategy Method)		
Game Method	- 2.314	-3.289
	(4.485)	(4.453)
Transparent Algorithm	10.720*	10.180*
	(4.755)	(4.759)
Black-Box Algorithm	10.940*	10.860*
	(4.682)	(4.694)
Female	- 7.241*	- 5.336
	(3.327)	(3.406)
Age	0.016	0.070
	(0.135)	(0.136)
Stated Risk Preferences		1.737**
		(0.661)
Familiar with the Trust Game		7.316*
		(3.384)
Failed the Comprehension Questions		2.719
		(7.811)
Constant	45.550***	45.670***
	(6.407)	(1.387)
N	799	795
Psuedo R2	0.003	0.004
F-statistic	4.080	4.290

* $p < 0.05$. ** $p < 0.01$, *** $p < 0.001$

These findings suggest, contrary to our hypothesis, that people trust algorithms more than other humans. This result is independent of the information revealed about the algorithm (transparent versus black-box algorithm) or the elicitation method used in the human conditions (strategy versus game method). Hence, both 'Hypothesis 1.1' and 'Hypothesis 1.2' are not supported.

On the contrary, we conclude that the amount sent in the algorithmic treatments are statistically significantly higher than the amount sent in the human conditions, while there is no difference in the amount sent in the algorithmic treatments. Thus, we have:

Result 1.1: People trust algorithms more than other people.

Result 1.2: The information revealed about an algorithm does not affect the trust towards the algorithm.

By looking at figure 3-6, it is clear that the source of the differences in the central tendencies lie in the boundaries of giving nothing and giving everything. While 22.50% of the senders in the SM and 24% of the senders in GM chose to keep the entire endowment, only 14.50% in the TA and 15% in the BBA chose to do the same. In contrast, 27% of the senders in the TA and 25.50% of the senders in the BBA chose to send the entire endowment, compared to 18.50% of senders in both the SM and the GM.

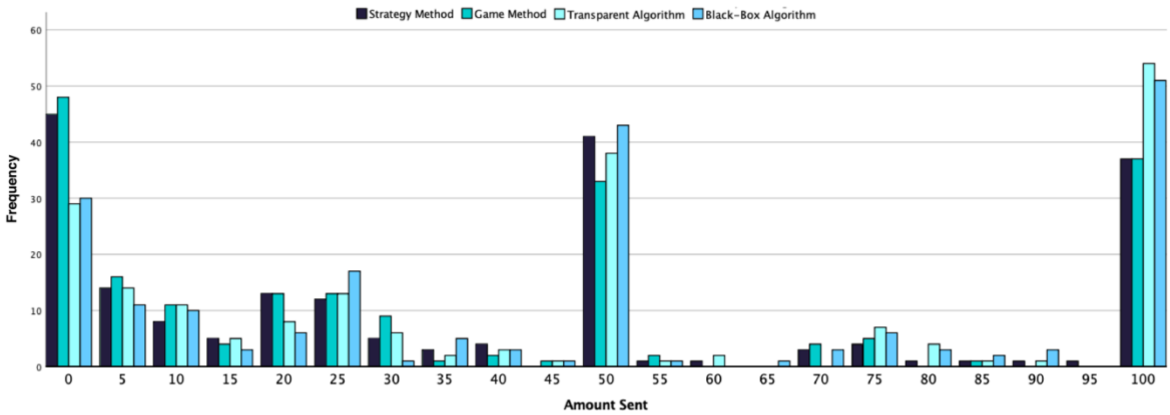


Figure 3-6: Comparison of amount sent across conditions

Figure 3-6 compares the distribution of amount sent by the senders in each condition. The participants had the option to send an amount between 0 and 100 cents in increments of five cents.

To address 'Hypothesis 1.3' we conducted a subsample analysis (Tobit regression, upper limit 100 cents) of both the transparent and the black-box algorithm treatment. By looking at 'Model 1 – 4' in table 3-5, we observe no significant effect of age in either of the algorithmic treatments (all $ps > 0.151$). Consequently, 'Hypothesis 1.3' is not supported and we conclude:

Result 1.3: Age does not affect people's trust in algorithms.

Table 3-5: Subsample analysis of the algorithmic treatments

Table 3-5 shows the outputs for the subsample analysis conducted on the algorithmic treatments. ‘Model 1’ and ‘Model 2’ are the outputs for the transparent algorithm treatment, while ‘Model 3’ and ‘Model 4’ show the outputs for the black-box algorithm treatment. Robust standard errors are reported in the parentheses below the estimated coefficients.

Dependent variable: Amount Sent				
	Tobit Regressions			
	Transparent Algorithm		Black Box Algorithm	
Regressor:	(1)	(2)	(3)	(4)
Female	- 7.070 (7.463)	- 6.524 (7.402)	- 3.265 (6.990)	- 0.987 (7.376)
Age	0.184 (0.315)	0.161 (0.316)	- 0.374 (0.259)	- 0.314 (0.268)
Stated Risk Preferences		0.866 (1.586)		1.203 (1.399)
Familiar with the Trust Game		15.610* (7.622)		5.943 (7.278)
Failed the Comprehension Questions		12.890 (21.870)		18.770 (25.520)
Constant	51.600*** (12.950)	40.680* (18.000)	70.800** (11.220)	58.22*** (16.270)
N	200	200	199	199
Psuedo R2	0.000	0.004	0.001	0.002
F-statistic	0.530	1.200	1.150	1.000

* p < 0.05. ** p < 0.01, *** p < 0.001

3.2.7. Discussion

‘Study 1’ addressed if people trust other people more than algorithms in one-shot interactions. The findings revealed considerable differences in participants behavior across conditions. Trustors who were matched with another human displayed less trust in their counterparts than those matched with an algorithm. This was independent of the elicitation methods employed in the human conditions and the information revealed about the algorithms in the algorithmic treatments. Hence, the results imply, contrary to our hypothesis, that people trust algorithms more than other humans. However, several scholars have noted that trustors might have other considerations in mind when deciding which strategy to employ in the trust game. Following, it is argued that the amount sent in the BDM trust game are affected by more than just trust, such as individual´s altruistic preferences (Cox, 2004) or betrayal aversion (Bohnet & Zeckhauser, 2004).

The idea of altruistic preferences suggests that people embed others' well-being into their own utility function. As such, altruistic preferences should, if present, manifest themselves in people's behavior. Specifically, participants should send more money in the human conditions compared to the algorithmic conditions—as altruistic participants would not send as much money to a machine as they would to a human. Yet, the results from the present study suggests the opposite. Indeed, trustors in both algorithmic conditions sent more of their endowment to their trustee compared to trustors in the human conditions. Thus, the presence of altruism, if any, mere reinforces the claim that people trust algorithms to a greater degree than humans.

According to the concept of betrayal aversion, people take more risk when the outcome of their decision is determined by chance rather than another person's trustworthiness—because people are averse to being betrayed (Bohnet & Zeckhauser, 2004). Consequently, Bohnet and Zeckhauser (2004, p. 470) hypothesized that it is 'fundamentally different to trust another person than to rely on a random device that offers the same outcome.' In the terminology of 'Study 1', parallels can be drawn between a random device and algorithms. Following, some may argue that the results from the present study could stem from betrayal aversion rather than trust. However, participants in the transparent algorithm treatment were informed that another person determined the outcome. Hence, our stance is that betrayal aversion cannot explain the observed differences between the human conditions and the transparent algorithm treatment. Moreover, as the behavior of participants in the algorithmic treatments were strikingly similar, we argue that the effect of betrayal aversion in the black-box algorithm treatment is, if present, not a dominating determinant of the participants behavior.

Another factor that might affect people's behavior in the trust game is individual's risk aversion. However, adding people's stated risk preferences as a control variable did not change the results. Hence, we argue that the presence of altruistic preferences, betrayal aversion or risk aversion cannot alone explain the observed differences. As such, we believe that 'Study 1' successfully elicited trusting behavior, and that the observed differences can, at least to a certain extent, be explained by the fact that people trust algorithms more than other people in the context of the trust game.

3.3. Study 2 – Judge Advisor System

3.3.1. Experimental design

To address the second research question ‘Do people rely more on financial advisors or robo-advisors?’ we built upon the existing experimental framework in the decision-making and advice-taking literature and manipulated the characteristics of the advisor. Similar to Önkal et. al. (2009), we employed the Judge-Advisor System (JAS) and asked subjects to forecast the weekly closing price for different stocks at two points in time: before and after receiving an advice. The source of the advice was told to be either a financial advisor (FA) or a robo-advisor (RA) depending on which condition the participants were assigned to.[20] The advice itself was identical in both conditions. If the participants’ final forecast was within 50 cents (either above or below) of the actual closing price of a randomly selected stock, they received a bonus of $1.00. If the final forecast was off by more than 50 cents (either above or below), no bonus was given.

Similar to previous studies (e.g. Logg et al., 2019; Prahl & Van Swol, 2017), participants were asked to rate their confidence in the accuracy of their prediction immediately after providing both their initial and final forecasts. This was done to see whether the participants’ confidence in their forecasts was affected by the received advice. Indeed, Sniezek and Van Swol (2001; Van Swol & Sniezek, 2005) found a strong relationship between participants’ reported confidence levels and trust in advice. As such, one would expect participants who utilize the advice to report a higher confidence level in their final forecast (post-advice forecast).

3.3.2. Measuring advice utilization

In line with previous literature (e.g. Önkal et al., 2009; Logg et al., 2019) we measured advice utilization by computing a positional measure (SHIFT) that captures the extent to which the participants adjust their forecast towards the advice given to them. The SHIFT variable is computed as follows:

$$\text{SHIFT} = \frac{(\text{Adjusted Forecast} - \text{Initial Forecast})}{(\text{Advisor Forecast} - \text{Initial Forecast})}$$

[20] Instructions and hyperlinks for each condition is attached in Appendix C.

A SHIFT-score of 0 indicates that the participant completely ignores the advice, while a score of 1.00 indicates that the participant discard their initial forecast and fully rely on the advice given by the advisor (Bonaccio & Dalal, 2006; Önkal et al., 2009). Note that if a participant's initial forecast is equal to the given advice, the SHIFT-score is undefined. Under such circumstances it is impossible to determine if the subject followed the advice. However, we did not find this to be a problem in our dataset as there were only three such cases (0.29% of all observations).

Another issue to address is SHIFT-scores above 1.00 or below 0. Negative SHIFT values occur when participants move away from the advice, while values above 1.00 occur when participants' final forecast overshoots the advice. Prior research has usually winsorized these observations by setting values greater than 1.00 to 1.00 and values lower than 0 to 0 (Gino & Moore, 2006; Logg et al., 2019). Although we agree that negative values indicate zero advice utilization, we argue that values greater than 1.00 does not reflect full advice utilization because the judge's final forecast overshoots and moves away from the given advice. Consequently, we adjusted these observations by subtracting twice the range between 1.00 and the original SHIFT, yielding a SHIFT below 1.00.[21] To exemplify, a SHIFT of 1.20 was adjusted to 0.80 and SHIFTs above 2.00 were adjusted to 0. Still, we recognize that a SHIFT of 0.80 and a SHIFT of 1.20 represent two fundamentally different behaviors. Therefore, all adjusted SHIFT-scores were assigned a dummy variable to keep them distinct from unadjusted SHIFTs in the analysis.[22]

3.3.3. Proceedings

Similar to 'Study 1', participants who satisfied the selection criteria (geographic location restricted to the US, minimum of 1,000 approved HITs, an approval rate of minimum 98% and a minimum age of 18) were recruited through Amazon Mechanical Turk. Individuals who participated in 'Study 1' were excluded to avoid biases based on recent experiences. For example, some participants were betrayed by their trustee in the trust game. This could

[21] Statistical tests of SHIFT-scores winsorized at 1.00 and 0 respectively [similar to Gino and Moore (2006) and Logg et al. (2019)] are attached in Appendix D.

[22] 6.10% of all the SHIFT-scores were above 1 and 1.24% were below 0.

potentially have affected participants' trust in us as experimenters, which in turn could have altered their behavior in this study.

Participants received a $1.00 show-up fee and were informed that they could earn a potential bonus depending the outcome of the experiment.[23] After giving their informed consent and passing the reCAPTCHA, they were randomly assigned to one of the two conditions (the financial advisor condition or the robo-advisor condition), before receiving some general information about the experimental session (see Appendix C). Next, all participants received the same instructions about the decision task to ensure common knowledge (see Appendix C). The participants were informed that they were to provide a four-week price forecast for three undisclosed stocks trading at the New York Stock Exchange based on some financial information (P/E, β, dividend and the 52-week range) and an interactive time series plot illustrating the 52-weekly closing prices for an undisclosed time period (see figure 3-7).[24] Both the companies name and the time period were not revealed to avoid potential biases (e.g. previous experiences with the stock or the 'mood' of the stock markets). Furthermore, we did not reveal that the participants would receive an advice and could revise their initial estimates. Finally, the participants were informed that the potential bonus would depend on the accuracy of their forecast.

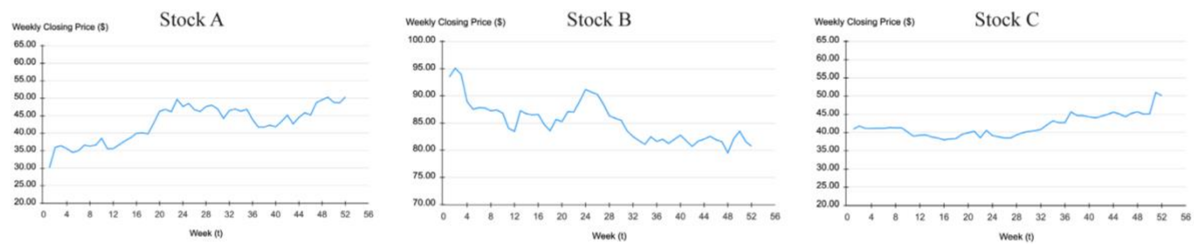


Figure 3-7: Stock Charts 'Study 2'

Figure 3-7 shows the interactive graphs the participants received for each forecasting task. Each graph illustrates 52-weekly closing prices. The participants were asked to forecast the weekly closing price in week 56.

After reading the instructions, the participants had to answer a set of comprehension question to ensure that they understood the task they were to perform. Participants who answered correctly proceeded to the decision phase of the experiment. Participants who failed the

[23] The HIT description for 'Study 2' is attached in Appendix C.

[24] The stocks belong to American Airlines, ExxonMobil and Oracle respectively. The data used was from the period June 2016 to July 2017.

questions after three attempts were forwarded to a page revealing the correct answers before they, too, continued to the decision phase.

The decision phase consisted of two parts. First, participants provided an initial forecast (text entry) and their confidence level (0-100 slider) for each stock. The order of the stocks was randomized to eliminate any systematic ordering biases. Next, the participants were introduced to their advisor (either a financial advisor or a robo-advisor) before revisiting the three stocks. For each stock, participants were reminded their initial forecast before receiving their advisor's recommendation. Participants were then asked to make their final forecast and report their corresponding confidence.

Once the participants had completed the forecasting tasks, they were forwarded to a post-questionnaire which included a manipulation check to see if the participants were aware of the source of the received advice. Upon completion, participants were able to see the results of the experiment and given a survey code to submit on MTurk in order to receive payment.

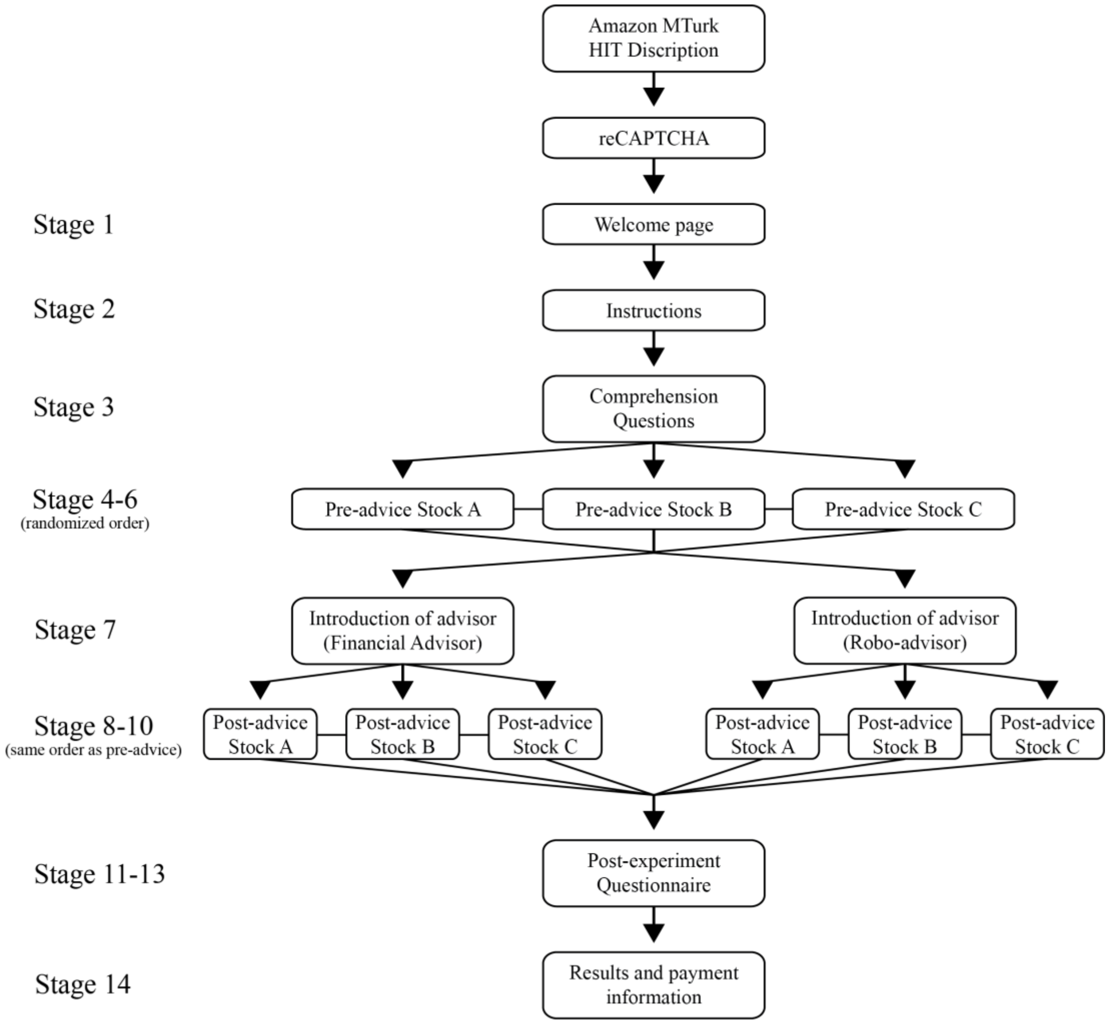


Figure 3-8: Flowchart 'Study 2'

Figure 3-8 illustrates the experimental flow for 'Study 2.' Participants were randomly assigned to receive an advice from either a financial advisor or a robo-advisor. Stage 4-6 were randomized to avoid ordering bias, and stage 8-10 were completed in the same order as stage 4-6.

3.3.4. Dropouts

Although measures were taken to prevent participation dropouts, 78 out of the 477 participants (16.35%) who entered the experimental session left before completing the experiment. As in 'Study 1', most of these participants (76.92%) left the experiment before the decision phase (see figure 3-9). Seven participants left once they entered the welcome page, five participants left when given the instructions and 48 participants left the experiment at the stage of the comprehension questions. Nine of the 18 dropouts that went on to the decision phase left without providing a forecast and six left before receiving the advice. Hence, only three of the 402 participants (0.75%) that provided an initial forecast for all three stocks left without completing the experiment.

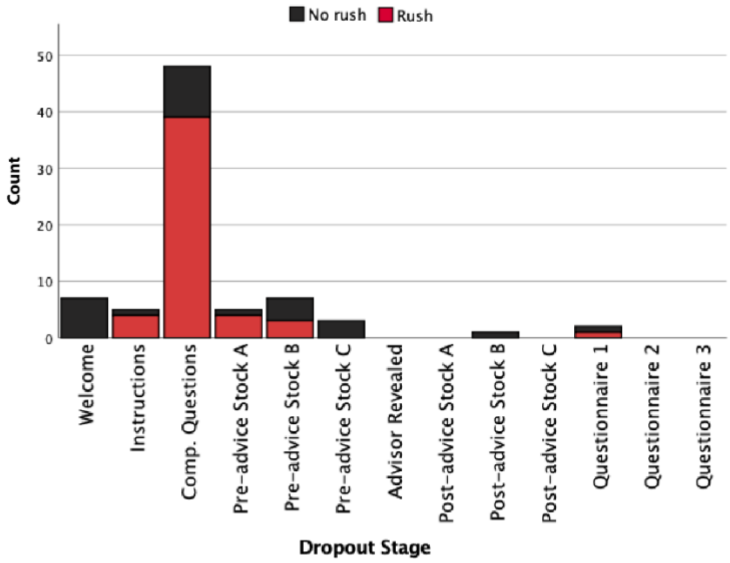


Figure 3-9: Participant dropout by stage

Figure 3-9 shows how many participants that dropped out at each stage of the experiment. The red shaded area illustrates how many participants that rushed through the stage (used 15 seconds or less to complete the stage).

A closer examination of the dropout data revealed that 8.97% of the dropouts left without receiving any crucial information about the decision task, while 65.38% of the dropouts rushed at least one stage before leaving the experiment.[25] From figure 3-9 it is clear that most of the dropouts left the experiment at the stage of the comprehension questions (48 of 78 dropouts or 61.54%). However, 81.25% of these (39 of 48 participants) rushed through at least one of the introductory pages (stage 1–2), which suggests that they did not read the instructions and were therefore unable to answer the questions successfully. Furthermore, 11 of the 15 participants who started the decision phase, but dropped out before receiving an advice, failed to answer the comprehension questions correctly. Five of these rushed through the instructions. Moreover, by comparing the number of dropouts in each condition, we found no statistical differences ($p = 0.534$, Chi-square test of homogeneity, see figure 3-10). As such, our stance is that selective dropouts did not lower the quality of the collected data.

[25] 'Rushed' means that the participant used 15 seconds or less to complete either the welcome page or the instructions.

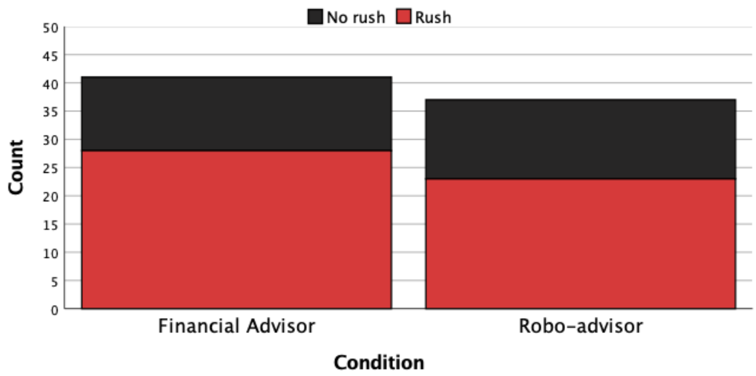


Figure 3-10: Participant dropout by condition

Figure 3-10 shows how many participants that dropped out in each condition. The red shaded area illustrates how many participants that rushed at least one stage during the experiment (used 15 seconds or less to complete at least one stage).

3.3.5. Sample

A total of 399 participants completed 'Study 2' (financial advisor, $n = 197$; robo-advisor, $n = 202$). 44 participants were removed from the final sample as they failed to answer the manipulation check by stating that the advice was given by an incorrect source, while five more were removed due to clear inconsistencies in the provided answers. This resulted in a final sample of 350 participants (financial advisor, $n = 169$; robo-advisor, $n = 181$).

As reported in table 3-6, the average age of the participants was 40.09 years old (SD 12.29) with the oldest participant being 81 and the youngest being 19. 174 of the participants were male (49.71%), while 175 participants (50.00%) where female. One participant (0.29%) reported a gender of 'other.' As in 'Study 1', participants with a bachelor's degree dominated the distribution (40.57%), followed by those with a high school diploma (24.00%) and technical training (17.14%). 58 participants had a master's degree (16.57%), while 6 had completed their PhD (1.71%). Moreover, 44.00 % of the participants had previously received financial advice. Of these, 71.43% had received advice from a financial advisor, 12.99% from a robo-advisor, while 15.58% had received financial advice from both. Finally, 7.43% of the participants that completed the experiment failed to answer at least one of the comprehension questions correctly and were assigned a dummy variable to use in the analysis.

Table 3-6: Distribution of age and gender 'Study 2'

Table 3-6 shows the distribution of age and gender for each condition, as well as for the whole sample. The financial advisor condition consisted of 169 participants, while the robo-advisor condition consisted of 181 participants. This yields a total sample of 350 participants.

		Total Sample	Financial Advisor	Robo-advisor
Age	Mean	40.09	41.18	39.08
	Std. Dev.	12.29	12.88	11.67
	Min	19	21	19
	Max	81	81	73
Gender	Male	49.71% (174)	53.25 % (90)	46.41% (84)
	Female	50.00% (175)	46.75 % (79)	53.04% (96)
	Other	0.29% (1)	- -	0.55% (1)

3.3.6. Predictions

Although research has found that statistical models and algorithms generally make more accurate predictions and forecasts than humans (e.g. Dawes, 1979; Meehl 1954), research on advice taking and decision-making often find that people frequently discount advice from non-human sources compared to advice given by a human (e.g. Dietvorst et al., 2015, 2018; Promberger et al., 2006; Önkal et al., 2009). However, other studies report the opposite (e.g. Logg et al., 2019; Madhavan & Wiegmann, 2007). Because of this ambiguity, Lee (2018) suggests that advice utilization may depend on the nature of the task being tested.

Prior research on financial decision-making suggests that humans are overconfident and possess cognitive biases that makes them bad investors. For example, people tend to seek information that confirms their existing beliefs (confirmation bias), overvalue the latest information available (recency bias) and neglect the role of probabilities (Ackert & Deaves, 2010). Moreover, people are loss averse and place a higher value on what they already own compared to what they do not (the endowment effect) (Ackert & Deaves, 2010). As a result of such biases, research show that investors trade too much (Odean, 1999), sell winners too quick and hold losers too long (Shefrin & Statman, 1985). Despite these findings, Önkal et al. (2009) found that people prefer human judgements over statistical recommendations in the domain of financial forecasting.

However, we argue that these results cannot be generalized as the sample used in the Önkal et al.'s study only included economics students with background from statistics and forecasting courses. Indeed, Prahl and Van Swol (2017 p. 694) note that such samples may lead to advice discounting and personal biases as subjects might overweigh their own financial literacy or be aware of the shortcomings of statistical models. Another point to note is that the study was conducted prior to the financial crisis (D. Önkal, personal communication, June 16, 2019). Hence, the results do not reflect the collapse of trust in financial professionals which may have changed people's behavior when facing financial advice.

In recent years, digital advisors that provide financial advice by utilizing algorithms have emerged as an alternative to traditional financial advisors. While research on human decision-making suggests that humans are bounded by overconfidence and cognitive biases, Fogg (2009) suggests that people trust computers to provide more unbiased information than humans, while Prahl and Van Swol (2017, p. 697) note that 'people generally expect automation to be perfect' (i.e. with an error rate of zero). Thus, we hypothesize:

> **Hypothesis 2.1:** People rely more on a financial advice emanating from a robo-advisor compared to a financial advisor.

Closely related to this hypothesis are the participants' reported confidence levels. As prior research has found a strong correlation between confidence and trust in advice (Sniezek & Van Swol, 2001; Van Swol, 2009; Van Swol & Sniezek, 2005), we predict that:

> **Hypothesis 2.2:** Participants who receive an advice from a robo-advisor will have a higher increase in their reported confidence levels compared to the participants in the financial advisor treatment.

Similar to 'Study 1', we predict that age will have an effect on the participants' behavior. Specifically, we expect reliance on robo-advice to decline with age. Indeed, several surveys have found that the use of robo-advisors is higher among younger investors (Financial Industry Regulatory Authority [FINRA], 2016; Lochner, Duenser & Reeson, 2017). Moreover, other remark that younger people are more comfortable with technology than older individuals (Fisch, Labouré & Turner, 2018). Thus, our last hypothesis is:

> **Hypothesis 2.3:** Younger people rely more on robo-advisors compared to older individuals.

3.3.7. Results

'Study 2' examined how people utilize financial advice depending on the source of the advice. We measured advice utilization by calculating a SHIFT-score that captures how individuals adjust their initial forecast towards the advice given to them. A higher SHIFT is associated with greater reliance on the advice.

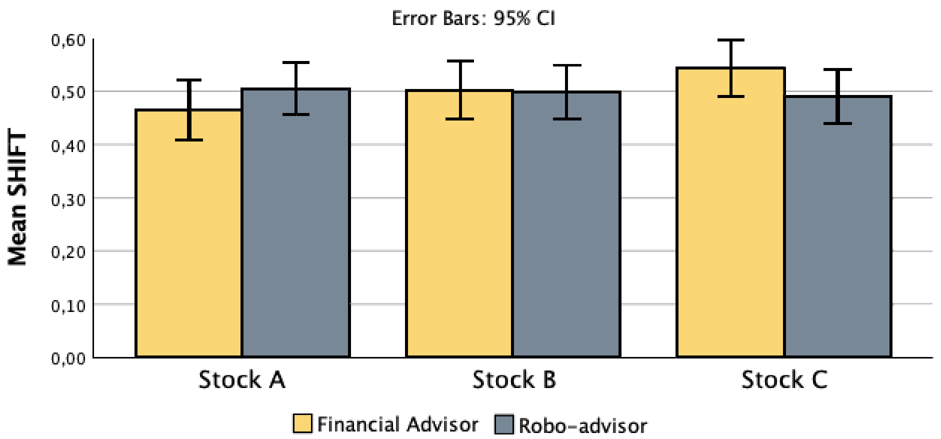


Figure 3-11: Average SHIFT-score by stock

Figure 3-11 illustrates the mean and the 95% confidence interval for each SHIFT-score (Stock A, Stock B and Stock C) by condition.

From figure 3-11, we see that both conditions have a mean SHIFT-score around 0.50 for all stocks. This implies that the participants averaged their own forecast with that of the advice. By comparing the SHIFT-score for of each stock across conditions, we observe no considerable differences (see table 3-7). The mean SHIFT-score for 'Stock A' is slightly higher in the robo-advisor treatment than in the financial advisor condition $(M = 0.505,\ SD = 0.337$ in the RA vs. $M = 0.464, SD = 0.363$ in the FA), while the opposite holds for 'Stock C' $(M = 0.490,\ SD = 0.337$ in the RA vs. $M = 0.542, SD = 0.351$ in the FA). However, the mean SHIFT of all forecasts combined is practically identical $(M = 0.498,\ SD = 0.339$ in the RA vs. $M = 0.503, SD = 0.359$ in the FA). This suggest that people rely equally on a financial advice given by a robo-advisor and a financial advisor, even though the advice is identical.

Table 3-7: Descriptive statistics SHIFT by condition

Table 3-7 shows the mean and standard deviation for each SHIFT-score, as well as the SHIFT-score for all stocks combined.

	Financial Advisor			Robo-advisor		
SHIFT	N	Mean	SD	N	Mean	SD
Stock A	166	0.464	0.363	181	0.505	0.337
Stock B	169	0.502	0.361	181	0.498	0.345
Stock C	169	0.542	0.351	181	0.490	0.337
Combined	504	0.503	0.359	543	0.498	0.339

A Hotelling's T^2 test of all three SHIFT-scores confirms the observations above (see Appendix D).[26] The differences between conditions on the combined SHIFT-scores were not statistically significant, $F(3, 343) = 1.335, p = 0.263, \text{Wilk's } \lambda = 0.988, \eta^2 = 0.012$. These results hold when controlling for gender, age, education level, perceived difficulty, confidence in the initial estimate, those who 'overshot' and those who failed the comprehension questions, $F(3, 325) = 1.388, p = 0.246, \text{Wilk's } \lambda = 0.987, \eta^2 = 0.013$ (one-way MANCOVA, see Appendix D).[27] As two assumptions were violated for both the Hotelling's T^2 test and the one-way MANCOVA (see Appendix D),[28] a Mann-Whitney U test was conducted for each SHIFT-score to check the robustness of the results. The tests showed no significant differences between the conditions, supporting the above findings (see table 3-8). Hence, 'Hypothesis 2.1' is not supported.[29] In fact, we conclude:

[26] The Hotelling's T^2 is a multivariate extension of the independent-samples t-test and a special case of the one-way MANOVA (multivariate analysis of variance) where the independent variable has only two groups. The Hotelling's T^2 is used to compare differences in means between two groups when you have more than one dependent variable (Laerd Statistics, 2017).

[27] The one-way MANCOVA is an extension of the one-way MANOVA where you can add control variables known as covariates (Leech, Barret & Morgan, 2005).

[28] The Hotelling's T^2 test violated the assumptions of normality and homogeneity of variances, while the one-way MANCOVA violated the assumptions of normality and linearity between the covariates and the dependent variables. However, it is worth noting that MANOVAs (e.g. Hotelling's T2 and the MANCOVA) is quite robust to deviations from normality and violations of homogeneity of variance, especially with equal (or nearly equal) sample sizes (Bray & Maxwell, 1985; Glass, Pekham & Sanders, 1972; Leech et al., 2005; Mardia, 1971).

[29] SHIFT-scores winsorized at 1.00 and 0 yields the same results (see Appendix D).

Result 2.1: People rely equally on a financial advice given by a financial advisor and a robo-advisor.

Table 3-8: Mann-Whitney U tests SHIFT

Table 3-8 presents the output of the Mann-Whitney U tests that were conducted to test differences in SHIFT-scores across each condition.

	Financial Advisor		Robo-advisor		Financial Advisor vs. Robo-advisor		
	Mean Rank	Sum of ranks	Mean Rank	Sum of ranks	U-value	z-value	p-value
Stock A	168.88	28 034.00	178.70	32 344.00	14 173.00	- 0.916	0.360
Stock B	176.73	29 868.00	174.35	31 557.00	15 086.00	- 0.222	0.824
Stock C	184.12	31 116.60	167.45	30 308.50	13 837.50	- 1.545	0.122
Combined	528.88	266 553.50	519.47	282 074.50	134 378.50	- 0.505	0.613

* p < 0.05. ** p < 0.01, *** p < 0.001

By looking at figure 3-12, we see that participants in both conditions reported a higher confidence in their forecast after receiving advice (see also table 8-21 in Appendix E). However, the magnitude of the increase does not seem to differ across the two conditions. The mean pre- and post-confidence levels for all stocks combined were 50.83% $(SD = 24.83)$ and 58.09 $(SD = 25.15)$ in the financial advisor condition versus 52.76% $(SD = 21.97)$ and 59.67% $(SD = 22.07)$ in the robo-advisor treatment. This yields an average of 7.20 $(SD = 13.66)$ and 6.92 $(SD = 13.84)$ percentage point change in confidence from time 1 to time 2 (before and after receiving advice) for each group respectively (see also table 8-21 in Appendix E).

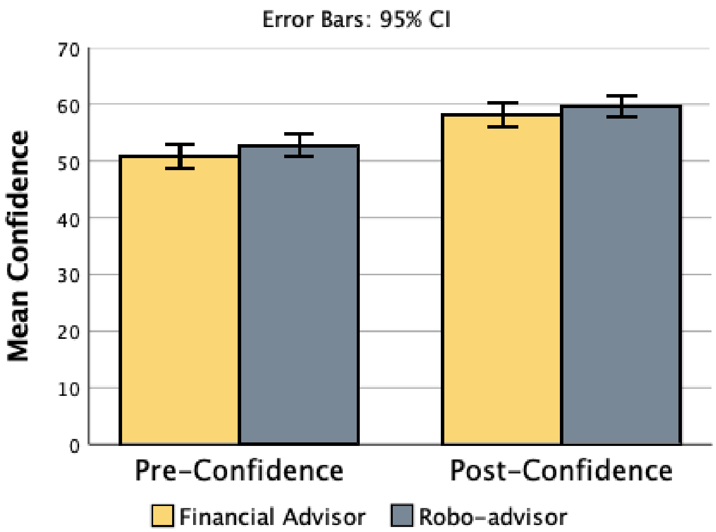


Figure 3-12: Mean confidence level by condition

Figure 3-12 illustrates the mean and the 95% confidence interval for the combined pre-advice and post-advice confidence levels (Stock A, Stock B and Stock C combined).

A two-way mixed MANOVA on all stocks was conducted to assess the statistical significance of these observations.[30] Receiving an advice had a positive effect on the participants' confidence levels, but no differences were found when comparing the change across the two conditions, as reflected in the main effect of advice on the combined dependent variables $F(3,345) = 47.343, p = 0.000, \text{Wilk's}\, \lambda = 0.708, \eta^2 = 0.292$, and the interaction between source and advice, $F(3,345) = 0.455, p = 0.714, \text{Wilk's}\, \lambda = 0.996, \eta^2 = 0.004$ (see Appendix D). Moreover, by examining the main effect of source (financial advisor or robo advisor), we found no difference in the participants' overall confidence level, $F(3,345) = 0.201, p = 0.896, \text{Wilk's}\, \lambda = 0.998, \eta^2 = 0.002$ (see Appendix D). This suggests that the participants were equally confident in their forecasts when advised by a financial advisor as they were when advised by a robo-advisor. However, as most of the assumptions for the two-way mixed MANOVA were violated (see Appendix D),[31] we calculated a change score (post-advice confidence minus pre-advice confidence) and conducted a Mann-Whitney U test for each stock, as well as for all stocks combined to check the robustness of the results. As reported in table 3-9, no significant difference was found in the change in confidence between the conditions.

Table 3-9: Mann-Whitney U tests change in confidence scores

Table 3-9 presents the output of the Mann-Whitney U tests that were conducted to test differences in the computed change-scores (post-advice confidence minus pre-advice confidence) for each stock across the conditions.

	Financial Advisor		Robo-advisor		Financial Advisor vs. Robo-advisor		
	Mean Rank	Sum of ranks	Mean Rank	Sum of ranks	U-value	z-value	p-value
Stock A	177.04	29 742.50	173.11	31 332.50	14 861.50	- 0.364	0.716
Stock B	169.48	32 782.50	181.12	32 782.50	14 277.50	- 1.077	0.282
Stock C	177.80	30 049.00	173.35	31 376.00	14 905.00	- 0.412	0.680
Combined	523.41	264 845.00	526.48	285 880.00	136 574.00	- 0.164	0.869

* p < 0.05. ** p < 0.01, *** p < 0.001

[30] A two-way mixed MANOVA is used to determine if there are differences between groups (between-subjects factor) and within subjects (within-subjects factor) when you have more than one continuous dependent variable (Laerd Statistics, 2018; Leech et al., 2005).

[31] The assumptions that were violated was the assumption of no multicollinearity, no univariate or multivariate outliers, multivariate normality and homogeneity of variances (see Appendix D).

These findings support the result that participants react equally to an advice given by a financial advisor and an advice given by a robo-advisor. Consequently, ‘Hypothesis 2.2’ is not supported and we have:

Result 2.2: People’s confidence in their predictions increase after receiving an advice, but the magnitude of the increase is independent of the source of the advice.

To address ‘Hypothesis 2.3’ and investigate whether age have a significant effect on advice utilization when the advice is given by a robo-advisor, we conducted a subsample analysis (OLS regression) of the robo-advisor treatment (see table 3-10).[32] By regressing the combined SHIFT-scores on gender, age, education level, forecasting task and those who overshot, we find no significant effect of age ($p = 0.109$). Controlling for the participants pre-confidence levels, the perceived difficulty of the forecasting task and those who failed the comprehension questions after three attempts does not change this result. This indicates that age does not affect the level of advice utilization when receiving advice from a robo-advisor. Hence, ‘Hypothesis 2.3’ is not supported and we have:

Result 2.3: People rely equally on an advice given by a robo-advisor regardless of age.

[32] Ordinary least squares (OLS) regression is a method used to predict the value of an unknown parameter based on the value of one or more independent variables. This is done by minimizing the sum of squared vertical distances between the observed values and the values predicted by linear approximation (Johannessen, Christoffersen & Tufte, 2011; Thrane, 2017).

Table 3-10: Subsample analysis of robo-advisor condition

Table 3-10 shows the outputs for the subsample analysis conducted on the robo-advisor condition. Robust standard errors are reported in the parentheses below the estimated coefficients.

Dependent variable: SHIFT (combined on all stocks)		
	OLS regressions	
Regressor:	(1)	(2)
Age	-0.002	-0.002
	(0.001)	(0.001)
Gender (ref: Male)		
Female	0.057	0.060*
	(0.030)	(0.030)
Other	0.305***	0.314***
	(0.061)	(0.063)
Education (ref: High School Diploma)		
Technical Training	-0.010	-0.010
	(0.044)	(0.045)
Bachelor's Degree	0.030	0.026
	(0.038)	(0.038)
Master's Degree	0.038	0.033
	(0.043)	(0.044)
Doctorate Degree	0.010	0.014
	(0.105)	(0.111)
Pre-condidence		-0.000
		(0.000)
Perceived Difficulty		0.002
		(0.009)
Forecasting Task (ref: Stock A)		
Stock B	-0.006	-0.006
	(0.036)	(0.036)
Stock C	-0.011	-0.011
	(0.035)	(0.036)
Failed the Comprehension Questions		0.047
		(0.053)
Overshot (Original SHIFT > 1.00)	0.142*	0.142*
	(0.060)	(0.061)
Constant	0.529***	0.526***
	(0.060)	(0.076)
N	543	543
R2	0.032	0.033
F-statistic	4.330	3.350

* p < 0.05. ** p < 0.01, *** p < 0.001

3.3.8. Discussion

'Study 2' examined if people rely more on advice emanating from a robo-advisor compared to a financial advisor. Previous literature comparing advice utilization from human sources versus non-human sources (e.g. algorithms or expert systems) often report that one source is more utilized than the other. However, our findings suggest that this is not the case when comparing financial advisors and robo-advisors in the domain of financial forecasting. Indeed, we find no statistical differences in advice utilization or participant's confidence in their forecasts depending on the source of the advice, which suggests that people rely equally on the two sources. This is in line with the results of Lee (2018) who found that people trust algorithmic and human decisions equally with tasks that requires mechanical skills.

Several scholars note that trust in financial advisors are affected by their perceived competence/ability, integrity and benevolence, which is the three components of trustworthiness suggested by Mayer et al. (1995). As the bonus in this experiment depended on the accuracy of the participants' forecasts, it is likely to believe that the perceived competence/ability of the advisor was the most salient factor when deciding how much to utilize the given advice. Indeed, Mayer et al. (1995) note that the importance of each component of trustworthiness differs according to the situation and argue that the 'domain of ability is specific because the trustee may be highly competent in some technical area, affording that person trust on tasks related to that area' (Mayer et al., 1995, p. 712). Moreover, a robust finding in the JAS literature is that the perceived expertise of the advisor increases the degree of advice utilization (Harvey & Fisher, 1997; Yaniv & Kleinberger, 2000; Jungermann & Fischer, 2005). As such, our result may suggest that financial advisors and robo-advisors are perceived to exhibit the same degree of competence, which can explain why participants display the same degree of advice utilization. However, this was not examined in the current research design and therefore needs to be investigated further to make a conclusion.

Interestingly, our participants tend to average their own forecast with that of the advisor, yielding a SHIFT of 0.50. This is in contrast to many other JAS studies which have found evidence of egocentric advice discounting (ref section 2.6.1). Closest to our experiment is the paper of Önkal et al. (2009), who found that people discounted advice from both a human expert and a statistical model when forecasting the future price of different stocks (SHIFTs of 0.39

and 0.28 respectively).[33] As Önkal et al.'s study was conducted prior to the financial crisis, it is interesting to note that our participants displayed a higher degree of advice utilization in general. However, several differences between the experiments can help explain the different results. First, the framing of the advisors (financial advisor vs. human expert and robo-advisor vs. statistical model) may have altered participants' behavior.[34] Second, the studies employed different reward structures, which have been found to influence the degree of advice utilization (Camerer & Hogarth, 1999; Sniezek & Van Swol, 2001).[35] Finally, Önkal et al. (2009) employed business and economics students enrolled in statistical and forecasting courses. The knowledge of, and previous experience with, financial forecasting may have reduced the perceived complexity of the task and made participants more confident in their own competence of stock price forecasting, which are two effects both associated with more advice discounting. In contrast, we recruited a diverse sample from Amazon Mechanical Turk. It is likely to believe that these participants had less financial and statistical literacy, which may help explain the higher SHIFT-scores in our experiment compared to the experiment of Önkal et al. (2009).

[33] Önkal et al. (2009) did not adjust SHIFT-scores above 1.00 or below 0.

[34] Framing the advisor as an expert or a statistical model rather than an advisor may explain some of the observed differences between the studies.

[35] While our participants could earn a $1.00 bonus depending on the accuracy of their forecasts, the subjects in Önkal et al.'s study received a fixed amount of extra credit points in their courses.

4. GENERAL DISCUSSION AND IMPLICATIONS

The main finding from 'Study 1' was that participants in both algorithm treatments sent a higher fraction of their endowment compared to participants in the human conditions, indicating that people trust algorithms more than other people. This result is quite robust as it is independent of the information revealed about the algorithm and the elicitation method used in the human conditions. Indeed, we found no differences when comparing the human conditions or algorithmic conditions. Hence, the different levels of trust seem to be a result of the characteristic of the trustee (algorithm versus human).

In light of this result, it is interesting that the participants in 'Study 2' relied equally on an advice given by a financial advisor compared to a robo-advisor. Although the experiments differ, it may seem like the general advantage algorithms have over humans in terms of trust cannot be particularized to the specific context of financial advisory. A possible explanation is offered by Lahance and Tang (2012), who found that trusting financial professionals is fundamentally different from trusting people. Related is the findings of Madhavan and Wiegmann (2007), which suggests that 'human novices' are trusted less than 'automated novices', while human experts are trusted more than automated experts. In fact, Logg et al. (2019) found that novice human advisers had lower credibility than automated advisers. The reverse held true when the human adviser was portrayed as an expert. Another potential explanation is that participants' in 'Study 1' might have based their decision on the beliefs about the dependent object's benevolence and integrity, while participants' in 'Study 2' used their subjective assessments of the advisor's credibility and competence when deciding how much to utilize their advice.

The results from the present studies have several implications. First, it illustrates the complexity of trust. Specifically, trust depends on the context and task at hand. This has both practical and theoretical implications. Although the results from 'Study 1' suggests that people are more willing to trust unknown algorithms more than unknown individuals, it is important for companies that want to implement and utilize algorithms in their business to know that trust in algorithms (and humans), largely depends on what kind of work or problem that is being addressed. Furthermore, it is important for trust researchers to be aware that their result may be affected by the experimental context and task under investigation as it may create a problem for the external validity of their research.

Second, the findings from ‘Study 2’ have practical implications for decisions-makers within the financial industry. As trust in financial advisors has been identified as to key determinant for seeking financial advice, it is essential for financial service providers to understand how trust in human financial advisors and robo-advisors differ as they are increasingly introducing digital financial advisors as part of their offerings. In addition, traditional providers of financial services need to be aware that people seem to be equally willing to rely on robo-advisors as financial advisors as the former are more accessible and cost-efficient. Providers of traditional financial services need to recognize a modernization of the industry and respond by developing an even better customer experience to remain competitive against the emerging computerized advisors.

Third, it is interesting to note that the transparency of an algorithm does not significantly change the level of trust. In everyday interactions with technology, underlying mechanisms of the algorithms involved are rarely exposed to users. Although there is an emerging trend emphasizing the importance of openness and disclosure of computational procedures, it will still be necessary for businesses to protect their products through proprietary software and undisclosed algorithmic processes. This result indicates that businesses can continue this practice without being punished in terms of lower trust towards their technologies.

Fourth, age does not seem to have an effect in either of the studies. Digitalization and modernization often focus on appealing to younger individuals. Yet, the results from our studies shows that trust in algorithms and reliance on financial advice from robo-advisors do not differ between generations. This is an important result for companies building their business on products based on technology and algorithms as it can affect how they present their products and target their consumers.

Finally, the findings from ‘Study 1’ have methodological implications. Specifically, it adds to the results from other behavioral experiments, which have found that framing effects (e.g. changing a few words in the instructions) and other experimental modifications can significantly alter participants’ behavior (Johnson & Mislin, 2011). Researchers need to be aware of this when generalizing their findings and evaluating the external validity of their results.

5. CONCLUSION

The aim of this thesis has been to contribute to the emerging research on trust in the interaction between humans and technology, both on a broader scale, and more specifically, in the domain of financial advisory. Section 1 addressed the background and motivation for the thesis, while section 2 provided an overview of previous trust literature. In particular, section 2 started with a discussion of the different conceptualizations, views and definitions of trust, followed by a brief review of various trust measurements and why people choose to trust others. Subsequently, we focused on trust in technology and the role of trust within the financial industry. The section ended with an introduction to the judge-advisor system and a presentation of previous research on advice utilization. Section 3 started with a presentation of Amazon Mechanical Turk and addressed the strengths and weaknesses of conducting online experiments. We then presented the experimental design, proceedings, predictions, results and a discussion for each study respectively. Section 4 provided a general discussion of the findings from the two studies, as well as some implications of the results.

In light of the discussions carried throughout this thesis, we conclude that the methods used to measure trust and reliance on advice provide valid results and are therefore suitable to answer the thesis' research questions presented in section 1. In particular, 'Study 1' provide evidence suggesting that people trust algorithms more than other people. This is in line with the findings of Wu et al. (2016). Despite this result, participants in 'Study 2' relied equally on a financial advice emanating from a human financial advisor compared to a robo-advisor.

The results from this thesis help contribute to the emerging focus on trust in technology, specifically by providing evidence suggesting that people trust algorithms more than other people on a general basis and that the transparency of an algorithm does not affect trust. Further, in the context of financial advisory, the thesis contributes with a finding that indicates no different reliance on financial advice from robo-advisors relative to traditional financial advisors. Yet, the latter result could be highly contingent on context, and the results may change if other domains than financial forecasting are investigated.

6. LIMITATIONS AND FUTURE RESEARCH

Like most research, this thesis has some limitations which provide opportunities for future studies. First, both 'Study 1' and 'Study 2' investigate one-shot interactions. However, several scholars have stated that trust tends to evolve with time (Lewicki, Tomlinson & Gillespie, 2006). Therefore, it would be interesting to see if our results could be replicated by including repeated interactions. Indeed, Mayer et al. (1995, p. 728) note that the outcome of previous interactions can indirectly affect trust 'through the perceptions of ability, benevolence and integrity at the next interaction.' For example, how would people's trust in algorithms change if the algorithm deviated from the trustor's expectations? In fact, previous research on non-human decisions aids have shown that people tend to punish algorithms more than people after seeing an algorithm err (Dietvorst et al., 2015, 2018). Future research should examine how this relates to the context of financial advisory and investigate how much people rely on an advice given by a robo-advisor after seeing the robo-advisor err.

A second limitation relates to the samples used in the experiments. Although MTurk offers a more diverse subject pool than those used in traditional laboratory experiments, it is still not a random sample of the general population (Mason & Suri, 2012; Paolacci & Chandler, 2014). Moreover, the geographical location of the participants was restricted to the United States. As previous studies have found that trust differs across countries and cultures (Johnson & Mislin, 2011), we encourage other researchers to attempt to replicate our findings in other countries and cultures—especially those with clear dissimilarities to the United States. For instance, samples from less developed countries may be confounded by variables such as lower education and less exposure to advanced technology—which may affect trust in technological artifacts like algorithms or robo-advisors.

Third, both studies were conducted online. Consequently, the interaction with the human counterparts in 'Study 1' and the advice given by the financial advisor in 'Study 2' occurred through a computer, not face-to-face. Riegelsberger, Sasse and McCarthy (2012) opine that the absence of physical interaction in online-environments may lead to higher uncertainty and lower trust due the loss of information resulting out of the lack of interpersonal cues. Moreover, Prahl and Van Swol (2017) remark that the reduced emotional and social cues may alter how the participants utilize an advice, while Önkal et al. (2009) note that in real life, advice recipients can question the human advisor and thereby get a deeper understanding of the underlying rationale of the advice. As both trust in people and advice utilization of human advice may be influenced by the form of social interaction, future research should investigate if the results from 'Study 1' and 'Study 2' hold when using face-to-face interactions.

Fourth, trust is context dependent (Mayer et al., 1995). As such, scholars across different academic disciplines should investigate how trust in people and trust in technology differ in their domain and academic field. For example, do patients rely more on a medical advice given by a doctor or a computerized advisor? This also relates to the domain of financial advisory. Indeed, this thesis were restricted to financial forecasting. However, few people seek financial advice to forecast the future price of financial assets. Hence, future research on financial advisory can focus on other tasks, such as investment decisions or debt counseling.

Fifth, there may be a limitation regarding the stakes involved in 'Study 2.' There is evidence that the size of the stakes does not influence people's behavior in experimental studies—with the exception of extremely high stakes (Amir et al., 2012; Raihani et al., 2013). Financial advisory may fall under this category, as investing in financial assets are associated with considerable investment risks. Unfortunately, it would not be feasible to run an experiment with stakes comparable to those present in the financial markets. One should note that the behavior in the real world may differ from the behavior observed in the experimental setting of this study.

Finally, this study measures trust at a single point in time. As technology constantly evolve, and algorithms and robo-advisors are in the adoption phase and still developing, trust towards them might change as they become more established. Future research should assess whether the results from this study hold true as people become more used to and familiar with algorithmic decision-aids. Furthermore, one should note that robo-advisors have never experienced a financial crisis. On the contrary, financial advisors have. As research has shown that algorithms are more heavily punished for making mistakes than humans (Dietvorst et al., 2015, 2018), it would be interesting to analyze what short and long-term effects an economic crisis would have on the results of a similar study.

7. REFERENCES

Accenture. (2017). *Financial Providers: Transforming distribution models for the envolving consumer*. Retrivied from https://www.accenture.com/us-en/insight-financial-services-distribution-marketing-consumer-study

Ackert, L. F. & Deaves, R. (2010). *Behavioral Finance: Psychology, Decision-Making and Markets*. South-Western Cengage Learning.

Amazon Mechanical Turk. (2019, 18. April). Qualifications and Worker Task Quality [Blog post]. Retrieved from https://bit.ly/31IOhuC

Amir, O., Rand, D. G. & Gal, Y. K. (2012). Economic games on the Internet: The effect of $1 stakes. *PLoS ONE*, 7(2). https://doi.org/10.1371/journal.pone.0031461

Arechar, A. A., Gächter, S. & Molleman, L. (2018). Conducting interactive experiments online. *Experimental Economics*, 21(1), 99-131. https://doi.org/10.1007/s10683-017-9527-2

Ashraf, N., Bohnes, I. & Piankov, N. (2006). Decomposing trust and trustworthiness. *Experimental Economics*, 9(3), 193-208. http://dx.doi.org/10.1007/s10683-006-9122-4

Bagheri, N. & Jamieson, G.A., (2004). Considering subjective trust and monitoring behavior in assessing automation-induced "complacency". In *Proceeding of the Human Performance, Situation Awareness and Automation Technology II Conference*, Daytona Beach, FL. Mahwah, NJ: L. Erlbaum Associates.

Banu, M. (2019). Why do we trust strangers? Social trust, moral reasoning and identity. *Annals of the University of Bucharest – Philosophy Series*, 67(2), 39-66. Retrieved from http://annals.ub-filosofie.ro/index.php/annals/article/view/278

Bauer, P. C. (2015). *Three essays on the concept of trust and its foundations* (Inaugural dissertation). Retrieved from https://boris.unibe.ch/69611/1/15bauer_p.pdf

Benbasat, I. & Wang, W. (2005). Trust in and adoption of online recommendation agents. *Journal of the Association for information systems*, 6(3), 72-101. Retrieved from https://aisel.aisnet.org/jais/vol6/iss3/4

Berg, J., Dickhaut, J. & McCabe, K. (1995). Trust, Reciprocity, and Social History. *Games and Economics Behavior*, *10*(1), 122-142. https://doi.org/10.1006/game.1995.1027

Berinsky, A. J., Huber, G. A. & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, 20(3), 351-368. https://doi.org/10.1093/pan/mpr057

Bichierri, C., Xiao, E. & Muldoon, R. (2011). Trustworthiness is a social norm, but trusting is not. *Politics, Philosophy & Economics*, 10(2), 170-187. https://doi.org/10.1177/1470594X10387260

Bohnet, I. & Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior & Organization*, 55(4), 467-484. https://doi.org/10.1016/j.jebo.2003.11.004

Bonaccio, S. & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes*, 101(2), 127-151. https://doi.org/10.1016/j.obhdp.2006.07.001

Bray, J. H. & Maxwell, S. E. (1985). Multivariate analysis of variance. Newbury Park, CA: Sage Publications.

Brüllhart, M. & Usunier, J. (2012). Does the trust game measure trust? *Economics Letters*, 115(1), 20-23. https://doi.org/10.1016/j.econlet.2011.11.039

Burke, J. & Hung, A. (2016). *Trust and financial advice* (RAND Working Paper Series WR-1075).

Camerer, C. F. & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, 19(1-3), 7-42. https://doi.org/10.1023/A:1007850605129

Chabert, J.-L., Barbin, È., Borowczyk, J., Guillemot, M. & Michel-Pajus, A. (1999). *A History of Algorithms: From the Pebble to the Microchip*. (J.-L. Charbert, Ed.). Berlin, Heidelberg: Springer.

Coleman, J. S. (1990). *Foundations of social theory*. Cambridge, Mass: Belknap Press of Harvard University Press.

Cox, J. (2004). How to identify trust and reciprocity. *Games and Economic Behavior*, 46(2), 260-281. https://doi.org/10.1016/S0899-8256(03)00119-2

Das, S. (2016, May 3). The Origin and Evolution of Algorithms. *digit*. Retrieved from https://bit.ly/2Xpk1Wo

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34(7), 571-582. https://doi.org/10.1037/0003-066X.34.7.571

Dietvorst, B. J., Simmons, J. P. & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114-126. http://dx.doi.org/10.1037/xge0000033

Dietvorst, B. J., Simmons, J. P. & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can even slightly modify them. *Management Science*, 64(3), 1155-1170. https://doi.org/10.1287/mnsc.2016.2643

Dietz, G. & Den Hartog, D. N. (2006). Measuring trust inside organisations. *Personnel Review*, 35(5), 557-588. https://doi.org/10.1108/00483480610682299

Dijkstra, J. J., Liebrand, W. B. G. & Timminga, E. (1998). Persuasiveness of expert systems. *Behaviour & Information Technology*, 17(3), 155-163. https://doi.org/10.1080/014492998119526

Dinesen, P. T. & Bekkers, R. (2015). The foundations of individuals' generalized social trust: A review. *Oxford Scholarship Online*. https://doi.org/10.1093/oso/9780190630782.003.0005

Doney, P. M., Cannon, J. P. & Mullen, M. R. (1998). Understanding the influence of national culture on the development of trust. *The Academy of Management Review*, 23(3), 601-620. https://doi.org/10.2307/259297

Elster, J. (2007) *Explaining Social Behavior*, Cambridge: Cambridge University Press.

Engelmann, D. & Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *The American Economic Review*, 94(4), 857-869. Retrieved from http://www.jstor.org/stable/3592796

Ennew, C. & Sekhon, H. (2007). Measuring trust in financial services: the trust index. *Consumer Policy Review*, 17(2), 62-68. Retrieved from https://bit.ly/2Xl5Hyg

Evans, A. M. & Krueger, J. I. (2009). The Psychology (and Economics) of Trust. *Social and Personality Psychology Compass, 3*(6), 1003-1017. https://doi.org/10.1111/j.1751-9004.2009.00232.x

Fehr, E. (2009). On the economics and biology of trust. *Journal of the European Economic Association*, 7(2-3), 235-266. https://doi.org/10.1162/JEEA.2009.7.2-3.235

Fisch, J.E., Labouré, M. & Turner, J. A. (2018). *The emergence of the robo-advisor* (PRC WP2018-12). Retrieved from https://pensionresearchcouncil.wharton.upenn.edu/wp-content/uploads/2018/12/WP-2018-12-Fisch-et-al.pdf

Financial Industry Regulatory Authority (2016). *Investors in the United States 2016* Retrieved from https://bit.ly/2XVz0EB

Friedman, B., Khan, P. H., Jr. & Howe, D. C. (2000). Trust online. *Communications of the ACM*, 43(12), 34-40. https://doi.org/10.1145/355112.355120

Fogg, B. J. & Tseng, H. (1999). The elements of computer credibility. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 80–87. Pittsburgh, PA. https://doi.org/10.1145/302979.303001

Fogg, B. J. (2009). A behavior model for persuasive design. In *Proceedings of the 4th International Conference on Persuasive Technology*. Claremont, CA. https://doi.org/10.1145/1541948.1541999

Gambetta, D. (1988). Can we trust trust? In D. Gambetta (ed.), *Trust: Making and breaking cooperative relations* (pp. 213-237). Cambridge, Massachusetts: Basil Blackwell.

Gino, F. & Moore, D. A. (2006). Effects of task difficulty on use of advice. *Journal of Behavioral Decision Making*, 20(1), 21-35. https://doi.org/10.1002/bdm.539

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A. & Soutter, C. L. (2000). Measuring trust. *The Quarterly Journal of Economics*, 115(3), 811-846. https://doi.org/10.1162/003355300554926

Glass, G., Peckham, P. & Sanders, J. (1972). Consequences of failure to meet assumptions underlying the fixed analysis of variances and covariance. *Review of Educational Research*, 42(3). https://doi.org/10.3102/00346543042003237

Google. (n.d.). reCAPTCHA. Retrieved from https://www.google.com/recaptcha/intro/v3.html

Guiso, L., Sapienza, P. & Zingales, L. (2008). Trusting the stock market. *The Journal of Finance*, 63(6), 2557-2600. https://doi.org/10.1111/j.1540-6261.2008.01408.x

Guiso, L. (2010*). A trust-driven financial crisis. Implications for the future of financial markets* (ECO Working Paper 2010/07). Retrieved from http://cadmus.eui.eu/handle/1814/13657

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

Hardin, R. (2002) *Trust & trustworthiness*. New York, NY: Russel Sage Foundation.

Hardin, R. (2006) *Trust*. Cambridge: Polity Press.

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

Heath, C. & Gonzalez, R. (1995). Interaction with others increases decision confidence but not decision quality: Evidence against information collection views of interactive decision-making. *Organizational Behavior and Human Decision Processes*, 61(3), 305–326. https://doi.org/10.1006/obhd.1995.1024

Hodge, F. D., Mendoza, K. I. & Sinha, R. K. (2018). The effect of humanizing robo-advisors on investor judgements. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3158004

Houser, D., Schunk, D. & Winter, J. (2010). Distinguishing trust from risk: An anatomy of the investment game. *Journal of Economic Behavior & Organization*, 74(1-2), 72-81. https://doi.org/10.1016/j.jebo.2010.01.002

Jilke, S., Van Ryzin, G. G. & Van de Walle, S. (2016) Responses to decline in marketized public services: An experimental evaluation of choice overload. *Journal of Public Administration Research and Theory*, 26(3), 421-432. https://doi.org/10.1093/jopart/muv021

Johannessen, A., Christoffersen, L. & Tufte, P. (2011). *Forskningsmetode for økonomisk-administrative fag* (3rd ed.). Oslo: Abstrakt.

Johnson, N. D. & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, 32(5), 865-889. https://doi.org/10.1016/j.joep.2011.05.007

Jungermann, H. & Fischer, K. (2005). Using expertise and experience for giving and taking advice. In T Betsch & S. Haberstroh (Eds.), *The Routines of Decision Making* (pp. 157-173). Mahwah, NJ: Lawrence Erlbaum Associates.

Karlan, D. S. (2005). Using experimental economics to measure social capital and predict financial decisions. *The American Economic Review*, 95(5), 1688-1699. Retrieved from https://www.jstor.org/stable/4132772?seq=1#metadata_info_tab_contents

Kaufmann, W. & Tummers, L. (2017). The negative effect of red tape on procedural satisfaction. *Public Management Review*, 19(9), 1311-1327. https://doi.org/10.1080/14719037.2016.1210907

Knights, D., Noble, F., Vurdubakis, T. & Willmott, H. (2001). Chasing shadows: control, virtuality and the production of trust. *Organizational studies*, 22(2), 311-336. https://doi.org/10.1177/0170840601222006

Kehl, D., Guo, P. & Kessler, S. (2017). Algorithms in the criminal justice system: Assessing the use of risk assessments in sentencing. Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School. Retrieved from https://dash.harvard.edu/handle/1/33746041

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

KPMG. (2019). *The pulse of fintech 2018*. Retrieved from https://home.kpmg/xx/en/home/insights/2019/01/pulse-of-fintech-h2-2018.html

Laerd Statistics. (2015). Mann-Whitney test using SPSS Statistics. *Statistical tutorials and software guides*. Retrieved from https://statistics.laerd.com/

Laerd Statistics. (2017). Hotelling's T^2 using SPSS Statistics. *Statistical tutorials and software guides*. Retrieved from https://statistics.laerd.com/

Laerd Statistics. (2018). Statistical test selector: Two-way mixed MANOVA. *Statistical tutorials and software guides*. Retrieved from https://statistics.laerd.com/premium/sts/groups-mixed-bw-c-no-cov-yes.php

Lahance, M. & Tang, N. (2012). Financial advice and trust. *Financial Services Review*, 21(3), 209-226. Retrieved from https://www.econbiz.de/Record/financial-advice-and-trust-lachance-marie-eve/10009714817

Lankton, N. K., McKnight, D. H. & Tripp, J. (2015). Technology, humanness and trust: Rethinking trust in technology. *Journal of the Association of Information Systems*, 16(10), 880-918. https://aisel.aisnet.org/jais/vol16/iss10/1

Lee J. D. & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1), 50-80. https://doi.org/10.1518/hfes.46.1.50_30392

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society, 5*(1), 1-16. https://doi.org/10.1177/2053951718756684r

Leech, N. L., Barrett, K. C. & Morgan, G. A. (2005). *SPSS for intermediate statistics: Use and interpretation* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Lewicki, R. J. & Brinsfield, C. (2011). Measuring trust beliefs and behaviours. In F. Lyon, G. Möllering & M. N. K. Saunders (Eds.), *Handbook of Research Methods on Trust* (pp. 1-15). Retrieved from https://www-dawsonera-com.ezproxy.uis.no/abstract/9780857932013

Lewicki, R. J., Tomlinson, E. C. & Gillespie, N. (2006). Models of interpersonal trust development: Theoretical approaches, empirical evidence, and future directions. *Journal of Management*, 32(6), 991-1022. https://doi.org/10.1177/0149206306294405

Lioness Lab (2018). *Lioness Lab: Reference manual*. Retrieved from https://lioness-doc.readthedocs.io/en/latest/040_reference_manual.html?fbclid=IwAR3O1H-vm5WQCwi_NnY_l7Qbw7FVjBvBmY4gQQy41VVVdofYKet_jQamMvU

Lochner, M., Duenser, A. & Reeson, A. (2017). *Automated financial advice for superannuation: A randomized controlled experimental test of engagement and motivation impacts* (Technical report). https://doi.org/10.4225/08/598372b89ffa4

Logg, J. M., Minson, J. A. & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes, 151*(2019), 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Luhmann, N. (1979). Trust and power. Chichester, Toronto: Wiley.

Lyon, F., Möllering, G. & Saunders, M. N. K. (2011). Introduction: The variety of methods for the multi-faceted phenomenon of trust. In F. Lyon, G. Möllering & M. N. K. Saunders (Eds.), *Handbook of Research Methods on Trust* (pp. 1-15). Retrieved from https://www-dawsonera-com.ezproxy.uis.no/abstract/9780857932013

Madamba, A. & Utkus, S. (2017). *Trust and financial advice*. Retrieved from https://vgi.vg/2ETib8c

Madhavan, P. & Wiegmann D.A. (2007). Effects of information source, pedigree, and reliability on operator interaction with decision support systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 49*(5), 773-785. https://doi.org/10.1518/001872007X230154

Mardia, K.V. (1971). The effect of non-normality on some multivariate tests and robustness to nonnormality in the linear model. *Biometrika,* 58(1). 105-121. https://doi.org/10.2307/2334321

Mason, W. & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, 44(1), 1-23. https://doi.org/10.3758/s13428-011-0124-6

Mayer, C. R., Davis J. H. & Schoorman D. F. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709-734. https://www.doi.org/10.2307/258792

McCabe, K., Houser, D., Ryan, L., Smith, V. & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 98(20), 11832–11835. https://doi.org/10.1073/pnas.211415698

McKnight, D. H. & Chervany, N. L. (2001). Trust and distrust definitions one bite at a time. In R. Falcone, M. Singh, & Y.-H. Tan (Eds.), *Trust in cyber-societies*. Berlin: Springer-Verlag.

McKnight, D. H., Carter, M., Thatcher J. B. & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems*, 2(2), 12-32. https://doi.org/10.1145/1985347.1985353

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis, University of Minnesota Press https://doi.org/10.1037/11281-000

Millar, M. (2012, November 9). Beating the recruitment machines. *BBC News*. Retrieved from https://www.bbc.com/news/business-20255387

Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies*, 27(5-6), 527-539. https://doi.org/10.1016/S0020-7373(87)80013-5

Naef, M. & Schupp, J. (2009). Measuring trust: Experiments and surveys in contrast and combination (IZA Discussion Paper no. 4087). http://dx.doi.org/10.2139/ssrn.1367375

Odean, T. (1999). Do investors trade too much? *American Economic Review*, 89(5), 1279-1298. Retrieved from http://www.jstor.org.ezproxy.uis.no/stable/117058

Önkal, D., Goodwin, P., Thomson, M., Gönül, S. & Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. Journal of Behavioral Decision Making, 22(4) 390-409. https://doi.org/10.1002/bdm.637

Paolacci, G. & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, 23(3), 184-188. https://doi.org/10.1177/0963721414531598

Prahl, A. & Van Swol, L. (2017). Understanding algorithm aversion: When is advice from automation discounted? *Journal of Forecasting*, 36(6), 691-702. https://doi.org/10.1002/for.2464

Promberger, M. & Baron, J. (2006). Do patients trust computers? *Journal of Behavioral Decision Making*, *19*(5), 455–468. https://doi.org/10.1002/bdm.542

Raihani, N. J., Mace, R. & Lamba, S. (2013). The effect of $1, $5 and $10 stakes in an online dictator game. *PLoS ONE*, 8(8). https://doi.org/10.1371/journal.pone.0073131

Reeves, B. & Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. New York, NY: Cambridge University Press.

Reichheld, F. F. & Schefter, P. (2000). E-loyalty: Your secret weapon on the web. *Harvard Business Review*, Jul-Aug, 105-113. Retrieved from https://hbr.org/2000/07/e-loyalty-your-secret-weapon-on-the-web

Reiersen, J. (2017). Trust as belief or behavior? *Review of Social Economy*, 75(4), 434-453. https://doi.org/10.1080/00346764.2016.1269939

Reiersen, J. (2019). Drivers of trust and trustworthiness. *International Journal of Social Economics*, 46(1), 2-17. https://doi.org/10.1108/IJSE-01-2018-0025

Riegelsberger, J., Sasse, M. A. & McCarthy, J. D. (2012). Trust in mediated interactions. In A Joinson, K. McKenna, T. Postmes & U. Reips (Eds.), *Oxford Handbook of Internet Psychology*. https://doi.org/10.1093/oxfordhb/9780199561803.013.0005

Robinson, S. L. (1996). Trust and breach of the psychological contract. *Administrative Science Quarterly*, 41(4), 574-599. https://doi.org/10.2307/2393868

Rotter, B. J. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, 35(4), 651-665. https://doi.org/10.1111/j.1467-6494.1967.tb01454.x

Rotter, B. J. (1971). Generalized expectancies for interpersonal trust. *American Psychologist*, 26(5), 443-452. http://dx.doi.org/10.1037/h0031464

Rousseau, D., Sitkin, S., Burt, R. & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*(3), 393–404. https://doi.org/10.5465/amr.1998.926617

Sapienza, P., Toldra-Simats, A. & Zingales, L. (2013). Understanding trust. *The Economic Journal*, 123(573), 1313-1332. https://doi.org/10.1111/ecoj.12036

Sapienza, P. & Zingales, L. (2012). A trust crisis. *International Review of Finance*, 12(2), 123-131. https://doi.org/10.1111/j.1468-2443.2012.01152.x

Schaefer, K. E. (2013). *The perception and measurement of human-robot trust* (Doctoral dissertation). Retrieved from https://stars.library.ucf.edu/etd/2688/

Schechter, L. (2007). Traditional trust measurement and the risk confound: An experiment in rural Paraguay. *Journal of Economic Behavior & Organization*, 62(2), 272-292. https://doi.org/10.1016/j.jebo.2005.03.006

Schrah, G. E., Dalal, R. S. & Sniezek, J. A. (2006). No decision-maker is an island: Integrating expert advice with information acquisition. *Journal of Behavioral Decision Making*, 19(1), 43-60. https://doi.org/10.1002/bdm.514

Science and Technology Committee. (2018). *4th Report - Algorithms in decision-making* (HC351). Retrieved from https://bit.ly/2GjhQu8

Shapiro, S. P. (1987). The social control of interpersonal trust. *American Journal of Sociology*, 93(3), 623-658. https://doi.org/10.1086/228791

Shefrin, H. & Statman, M. (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *The Journal of Finance*, 40(3), 777-790. https://doi.org/10.2307/2327802

Shneiderman, B. (2000). Designing trust into online experiences. *Communications of the ACM*, 43(12), 57-59. https://doi.org/10.1145/355112.355124

Sniezek, J. A. & Buckley, T. (1995). Cueing and cognitive conflict in Judge-Advisor decision making. *Organizational Behavior and Human Decision Processes*, 62(2), 159-174. http://dx.doi.org/10.1006/obhd.1995.1040

Sniezek, J. A. & Van Swol L. M. (2001). Trust, confidence, and expertise in a judge-advisor system. *Organizational Behavior and Human Decision Processes*, 84(2), 288-307. https://doi.org/10.1006/obhd.2000.2926

Sunikka, A., Peura-Kapanen, L. & Raijas, A. (2010). Empirical investigation into the multi-faceted trust in the wealth management context. *International Journal of Bank Marketing*, 28(1), 65-81. https://doi.org/10.1108/02652321011013599

Sztompka, P. (1999). *Trust: A Sociological Theory*. Cambridge, UK: Cambridge University Press.

Tabachnick, B. G. & Fidell, L. S. (2012). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.

The Economist. (2018, June 21). How an algorithm may decide your career. *The Economist*. Retrieved from https://www.economist.com/business/2018/06/21/how-an-algorithm-may-decide-your-career

Thielmann, I. & Hilbig, B. E. (2015). The traits one can trust: Dissecting reciprocity and kindness as determinants of trustworthy behavior. *Personality and Social Psychology Bulletin*, 41(11), 1523-1536. https://doi.org/10.1177/0146167215600530

Thrane, C. (2017). *Regresjonsanalyse: En praktisk tilnærming*. Oslo: Cappelen Damm Akademisk.

Treanor, J. (2014, October 26). Digital revolution presents banks with more change in 10 years than last 200. *The Guardian*. Retrieved from https://bit.ly/2Rr0len

UC Berkley Committee for Protection of Human Subjects. (2018). Mechanical turk (MTurk) for online. Retrieved from https://bit.ly/2XzbvEw

UCLA Institute for Digital Research & Education. (n.d.). Tobit analysis: Stata data analysis examples. Retrieved from https://stats.idre.ucla.edu/stata/dae/tobit-analysis/

Van Thiel, D., Goedee, J. & Reijnders, W. (2008). *Riding the waves. The future of banking starts now*. Amsterdam: Pearson Financial Times Prentice Hall.

Van Swol, L. M. & Sniezek, J. A. (2005). Factors affecting the acceptance of expert advice. *The British Journal of Social Psychology*, 44(Pt 3), 443-461. https://doi.org/10.1348/014466604X17092

Van Swol, L. M. (2009). The effects of confidence and advisor motives on advice utilization.

Communication Research, 36(6), 857-873. https://doi.org/10.1177/0093650209346803

Walterbusch, M., Gräuler, M. & Teuteberg, F. (2014). *How trust is defined: A qualitative and quantative analysis of scientific literature*. Presented at the 20th Americas Conference on Information Systems (AMCIS), Savannah, Georgia. Retrieved from https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1142&context=amcis2014

Wu, J., Paeng, E., Linder, K., Valdesolo, P. & Boerkoel, J. C. (2016). Trust and Cooperation in Human-Robot Decision Making. In The *2016 AAAI Fall Symposium Series: Artificial Intelligence for Human-Robot Interaction,* 110-116. Retrieved from https://aaai.org/ocs/index.php/FSS/FSS16/paper/view/14118

Wu, K., Zhao, Y., Zhu, Q., Tan, X. & Zheng, H. (2011). A meta-analysis of the impact of trust on technology acceptance model: Investigation of moderating influence of subject and context type. *International Journal of Information Management*, 31(6), 572-581. https://doi.org/10.1016/j.ijinfomgt.2011.03.004

Xu, J., Le, K., Deitermann, A. & Montague, E. (2014). How different types of users develop trust in technology: A qualitative analysis of the antecedents of active and passive user trust in a shared technology. *Applied Ergonomics*, 45(6), 1495-1503. https://doi.org/10.1016/j.apergo.2014.04.012

Yamagishi, T. & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion*, 18(2), 129-166. https://doi.org/10.1007/BF02249397

Yaniv, I. & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes*, 83(2), 260-281). https://doi.org/10.1006/obhd.2000.2909

Yeomans, M., Shah, A., Mullainathan, S. & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making*, 1-12. https://doi.org/10.1002/bdm.2118

Yousefi, N. & Nasiripour, A. (2015). A Proposed Model of E-Trust for Electronic Banking. *Management Science Letters*, 5 (11), 1029–1040. https://doi.org/10.5267/j.msl.2015.8.008

Zhou, H. & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology*, 111(4), 493-504. https://doi.org/10.1037/pspa0000056

8. APPENDICES

8.1. Appendix A: Instructions 'Study 1'

8.1.1. Mturk HIT description

8.1.1.1. Pre-experiment

Economics Experiment for Research Purposes

Key information about the HIT:

- The HIT will take *approximately 15 minutes*.

Payment:

- You will be paid a $1.00 participation fee within three days upon completion of the HIT.
- You will earn a bonus between $1.00 and $4.00 which will be paid later.
- At the end of the experiment, you will receive a code to paste into the box below to receive your payment.

About the experiment:

- This HIT is an academic experiment on economic decision-making.
- The experiment consists of twenty decisions and a short questionnaire.
- Before the experiment starts you will have to read some instructions and answer some simple control questions.

Browser compatibility:

- The experiment will not work using Internet Explorer.
- To perform the experiment, you must have Javascript and cookies enabled in your browser.

Voluntary participation and confidentiality:

- Participation in this HIT is entirely voluntary.
- Your answers are confidential and will not be revealed to anyone other than the experimenters.
- Your participation in the experiment is anonymous. No one, including the experimenters, will know any personal information about you.
- To know more about how we handle your privacy, please click here.

Informed consent:

- By accepting to participate in this HIT you give us informed consent that we can use your answers in an anonymized form for research purposes only.

Other information:

- Your Worker ID will be retrieved automatically when you click the link to the external survey page. Your Worker ID will only be used to secure payment to the right account and to control that you have not participated in this study before.
- **Make sure to leave this window open as you complete the experiment.** When you are finished, you must return to this page to paste the code into the box below and submit the HIT.

8.1.1.2 Main Experiment

Economics Experiment for Research Purposes

Key information about the HIT:

- The HIT will take *approximately 15 minutes*.
- **Due to the nature of the experiment, we kindly ask you to start the experiment *immediately* after accepting this HIT and to complete the experiment in one go without interruptions.**

Payment:

- You will be paid a $1.00 participation fee within three days upon completion of the HIT.
- You can earn a bonus that will be paid later.
- At the end of the experiment, you will receive a code to paste into the box below to receive your payment.

About the experiment:

- This HIT is an academic experiment on economic decision-making.
- Before the experiment starts you will have to read some instructions and answer some simple control questions.
- The experiment consists of one decision-making task and a questionnaire containing nine questions.
- There might be some waiting during the experiment.

Browser compatibility:

- The experiment will not work using Internet Explorer.
- To perform the experiment, you must have Javascript and cookies enabled in your browser.

Voluntary participation and confidentiality:

- Participation in this HIT is entirely voluntary.
- Your answers are confidential and will not be revealed to anyone other than the experimenters.
- Your participation in the experiment is anonymous. No one, including the experimenters, will know any personal information about you.
- To know more about how we handle your privacy, please click here.

Informed consent:

- By accepting to participate in this HIT you give us informed consent that we can use your answers in an anonymized form for research purposes only.

Other information:

- Your Worker ID will be retrieved automatically when you click the link to the external survey page. Your Worker ID will only be used to secure payment to the right account and to control that you have not participated in this study before.
- **Make sure to leave this window open as you complete the experiment.** When you are finished, you must return to this page to paste the code into the box below and submit the HIT.

8.1.2. Welcome page – Stage 1

Welcome!

This HIT consists of two parts: (1) An economics experiment and (2) a questionnaire containing nine (9) questions. The HIT will take approximately 15 minutes to complete and must be completed in one go without interruptions. Both the experiment and the questionnaire must be completed to receive payment. Therefore, please only participate if you can dedicate at least 15 minutes of your time.

You will be paid a participation fee of $1 for completing this HIT. In addition, you will be able to earn a bonus depending on your decision and the outcome of the experiment. Upon completion, you will receive a code to collect your payment via MTurk. The $1 participation fee will be paid within three days upon completion of this HIT. The potential bonus will be paid later.

Before the experiment starts, you have to read some instructions. These instructions should be self-explanatory. However, if you have any questions or need clarifications, you should read back through the instructions. Once you have read the instructions, you will need to answer some simple control questions to make sure you have understood the task you are to perform. Please note that a summary of the instructions is available at every stage of the experiment.

During the experiment, please do not close this window or leave the HIT´s web pages in any other way. If you do close your browser or leave the experiment, you will not be able to re-enter, and we will not be able to pay you!

When you are ready to proceed, please click on the button below.

8.1.3. General instructions – Stage 2

Instructions

Thank you for accepting to participate in this experiment. Please read the instructions below carefully and make sure you understand the experiment.

This experiment is conducted in pairs of senders and responders. You will randomly be assigned to one of two roles: (1) a sender or (2) a responder. At the start of the experiment, both the sender and the responder are given 100 cents. The sender then gets the opportunity to send all, some or none of his/her money to the responder in increments of five (5) cents. The amount that is not sent is kept by the sender. The amount that is sent to the responder will be tripled. That is, if the sender chooses to send X cents to the responder, the responder receives 3*X cents. The responder then decides how much of this money to keep and how much to return to the sender. The responder can return all, some or none of the received money in increments of five (5) cents. The amount returned to the sender will not be tripled.

In summary, the payoffs from the experiment will be as follows:

Sender:	100¢ - Amount sent to the responder + Amount returned by the responder
Responder:	100¢ + 3*Amount received from the sender - Amount returned to the sender

Note that this payoff is in addition to the guaranteed participation fee for completing the HIT.

8.1.4. Specific instructions senders – Stage 3

8.1.4.1. Strategy Method – Baseline

Your Role

You have randomly been given the role of a sender. The responder you will be paired with is another Amazon Mechanical Turk worker, from now on called “Person B”. Both you and “Person B” have been given the same instructions about this experiment, but you are not completing the experiment at the same time. “Person B” has already given his/her response to whatever you choose to send to “Person B” in this experiment. You will not be given any more information about “Person B”.

On the next page you will be asked to answer some simple control questions to make sure you have understood the instructions and the task you are to perform. If you at that point still need some clarifications, you should look at the summary which is available at every stage of the experiment.

8.1.4.2. Game Method – Treatment 1

Your Role

You have randomly been given the role of a sender. The responder you will be paired with is another Amazon Mechanical Turk worker, from now on called “Person B”. Both you and “Person B” have been given the same instructions about the experiment and you are completing the experiment at the same time. You will not be given any more information about "Person B".

On the next page, you will be asked to answer some simple control questions to make sure you have understood the instructions and the task you are to perform. If you at that point still need some clarifications, you should look at the summary which is available at every stage of the experiment.

8.1.4.3. Transparent Algorithm – Treatment 2

Your Role

You have randomly been given the role of a sender. The responder you will be paired with is a pre-programmed algorithm from now on called “The Algorithm”. The algorithm was programmed by an Amazon Mechanical Turk worker, from now on called “The Programmer”. While programming, the programmer was given the same instructions as you about this experiment. The programmer then decided how the algorithm will act in response to whatever you choose to send to the algorithm in this experiment. You will not be given any more information about the algorithm or the person who programmed it.

On the next page, you will be asked to answer some simple control questions to make sure you have understood the instructions and the task you are to perform. If you at that point still need some clarifications, you should look at the summary which is available at every stage of the experiment.

8.1.4.4. Black-Box Algorithm – Treatment 3

Your Role

You have randomly been given the role of a sender. The responder you will be paired with is a pre-programmed algorithm from now on called “The Algorithm”. You will not be given any more information about the algorithm.

On the next page you will be asked to answer some simple control questions to make sure you have understood the instructions and the task you are to perform. If you at that point still need some clarifications, you should look at the summary which is available at every stage of the experiment.

8.1.5. Specific instructions responders – Stage 3

8.1.5.1. Strategy Method – Baseline

Your Role

In the future economics experiment, you are given the role of a responder. You will be paired with a sender which is another Amazon Mechanical Turk worker, from now on called "Person A". "Person A" will receive the same information as you about this experiment. "Person A" will also be informed that the responder is another Amazon Mechanical Worker that has already provided a response to whatever amount "Person A" chooses to send to the responder. "Person A" will not be given any more information about you.

Your task in this HIT is to answer 20 questions, one for each of the scenarios depending on what "Person A" chooses to send to you. Specifically, you will decide how much you shall return to "Person A" given "Person A´s" decision in the future experiment.

On the next page, you will be asked to answer some simple control questions to make sure you have understood the instructions and the task you are to perform. If you at that point still need some clarifications, you should look at the summary which is available at every stage of the HIT.

8.1.5.2. Game Method – Treatment 1

Your Role

You have randomly been given the role of a responder. The sender you will be paired with is another Amazon Mechanical Turk worker, from now on called "Person A". Both you and "Person A" have been given the same instructions about the experiment and you are completing the experiment at the same time. You will not be given any more information about "Person A".

On the next page, you will be asked to answer some simple control questions to make sure you have understood the instructions and the task you are to perform. If you at that point still need some clarifications, you should look at the summary which is available at every stage of the experiment.

8.1.5.3. Transparent Algorithm – Treatment 2

Your Role

In the future economics experiment, your algorithm will be given the role of a responder. The sender your algorithm will be paired with is another Amazon Mechanical Turk worker, from now on called “Person A”. “Person A” will receive the same information as you about the experiment, as well as how your algorithm was programmed. That is, “Person A” will be informed that another Amazon Mechanical Turk worker has decided how the algorithm will act in response to whatever “Person A” chooses to send to the algorithm. “Person A” will not be given any more information about you or your algorithm.

Your task in this HIT is to help program an algorithm. All you have to do is answer 20 questions, one for each of the scenarios depending on what “Person A” chooses to send to your algorithm. Specifically, you will decide how much your algorithm shall return to “Person A” given “Person A´s” decision in the future experiment.

On the next page, you will be asked to answer some simple control questions to make sure you have understood the instructions and the task you are to perform. If you at that point still need some clarifications, you should look at the summary which is available at every stage of the HIT.

8.1.5.4. Black-Box Algorithm – Treatment 3

Your Role

In the future economics experiment, your algorithm will be given the role of a responder. The sender your algorithm will be paired with is another Amazon Mechanical Turk worker, from now on called “Person A”. “Person A” will be informed that the responder is an algorithm, but he/she will not be given any information about you or how the algorithm was programmed.

Your task in this HIT is to help program an algorithm. All you have to do is answer 20 questions, one for each of the scenarios depending on what “Person A” chooses to send to your algorithm. Specifically, you will decide how much your algorithm shall return to “Person A” given “Person A´s” decision in the future experiment.

On the next page, you will be asked to answer some simple control questions to make sure you have understood the instructions and the task you are to perform. If you at that point still need some clarifications, you should look at the summary which is available at every stage of the HIT.

8.1.6. Summary General Instructions

SUMMARY

The table below summarizes the range of opportunities for the sender and the amount the responder receives following the sender´s decision.

Sent by the sender: (X)	The responder receives: (3*X)
0¢	0¢
5¢	15¢
10¢	30¢
15¢	45¢
20¢	60¢
25¢	75¢
30¢	90¢
35¢	105¢
40¢	120¢
45¢	135¢
50¢	150¢
55¢	165¢
60¢	180¢
65¢	205¢
70¢	210¢
75¢	225¢
80¢	240¢
85¢	255¢
90¢	270¢
95¢	285¢
100¢	300¢

8.1.7. Summary stage 3-5

SUMMARY

- You are given the role of a **XX**.
- You can choose to send all, some or none of the 100 cents.

The table below summarizes your opportunities and the amount "Person B"/the algorithm will receive following your decision.

Sent by the sender: (X)	The Person B"/the algorithm receives: (3*X)
0¢	0¢
5¢	15¢
10¢	30¢
15¢	45¢
20¢	60¢
25¢	75¢
30¢	90¢
35¢	105¢
40¢	120¢
45¢	135¢
50¢	150¢
55¢	165¢
60¢	180¢
65¢	205¢
70¢	210¢
75¢	225¢
80¢	240¢
85¢	255¢
90¢	270¢
95¢	285¢
100¢	300¢

8.1.8. Hyperlinks for the main experiment `Study 1`

The matching process in the game method treatment will not work as there is no server running to host the experiment which is required for the lobby function to work.

Strategy Method – Baseline - Sender

https://start.econexperiment.org/main/_beginParticipant.php?workerID=CG

Game Method – Treatment 1 - Sender

https://start.econexperiment.org/main/_beginParticipant.php?workerID=T1

Game Method – Treatment 1 - Responder

https://start.econexperiment.org/main/_beginParticipant.php?workerID=T1R

Transparent Algorithm – Treatment 2 - Sender

https://start.econexperiment.org/main/_beginParticipant.php?workerID=T2

Black-Box Algorithm – Treatment 3 - Sender

https://start.econexperiment.org/main/_beginParticipant.php?workerID=T3

8.1.9. Hyperlinks for the pre experiment `Study 1`

Strategy Method – Baseline - Responder

https://start.econexperiment.org/5/_beginParticipant.php

Transparent Algorithm – Treatment 2 - Responder

https://start.econexperiment.org/4/_beginParticipant.php

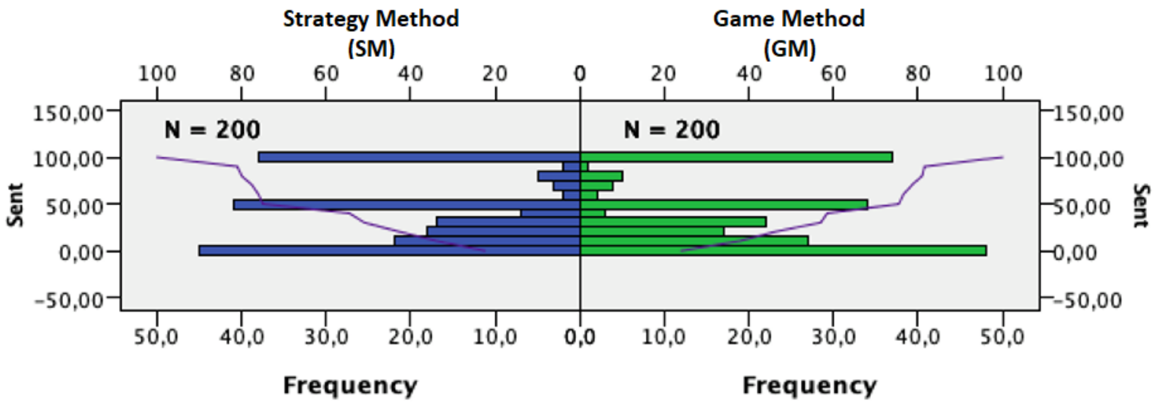
Black-Box Algorithm – Treatment 3 - Responder

https://start.econexperiment.org/3/_beginParticipant.php

8.2. Appendix B: Statistical tests ‘Study 1

8.2.1. Independent-Samples Kolmogorov-Smirnov Test

8.2.1.1. Strategy Method versus Game Method

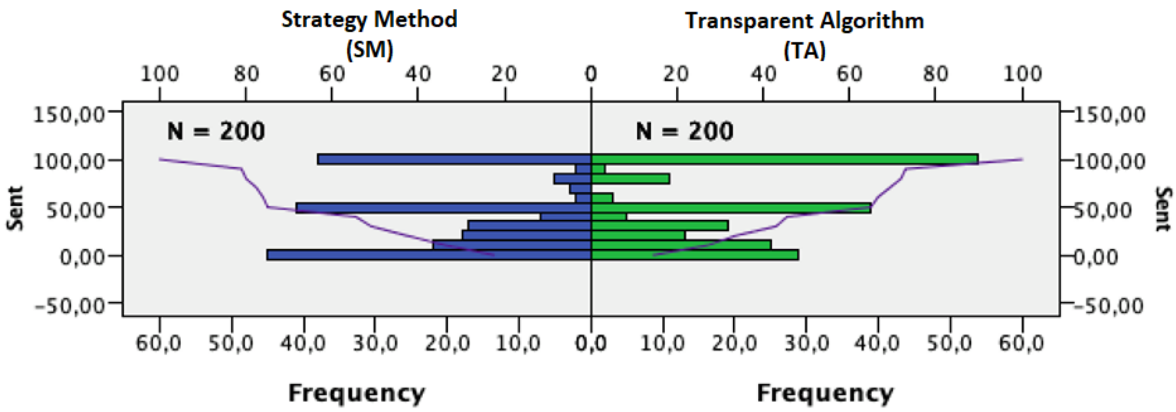


Total N		400
Most Extreme Differences	**Absolute**	,060
	Positive	,060
	Negative	,000
Test Statistic		,600
Asymptotic Sig. (2-sided test)		,864

Figure 8-1: Strategy Method versus Game Method

Figure 8-1 shows the output of the independent-samples Kolmogorov-Smirnov test which compares the distribution of the amount sent in the strategy method condition and the game method treatment.

8.2.1.2. Strategy Method versus Transparent Algorithm

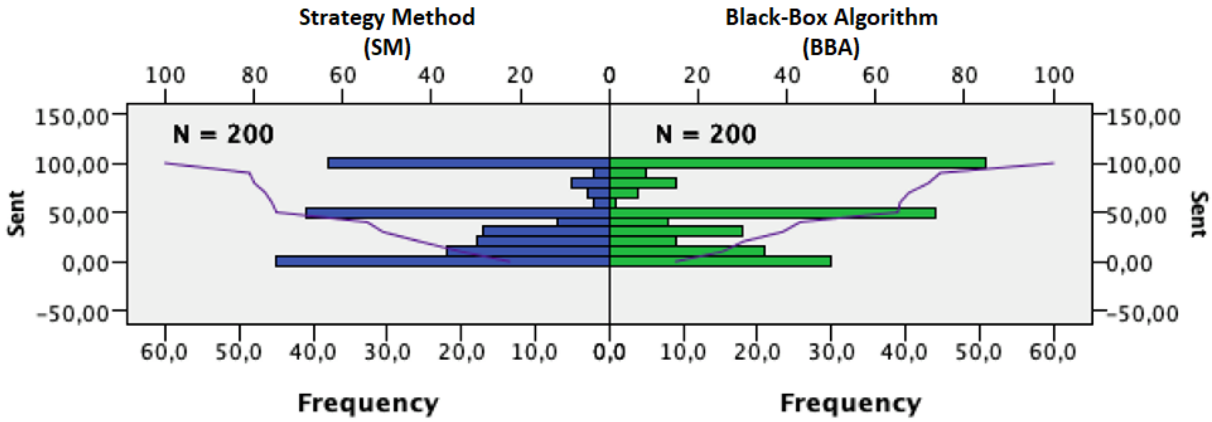


Total N		400
Most Extreme Differences	Absolute	,110
	Positive	,000
	Negative	-,110
Test Statistic		1,100
Asymptotic Sig. (2-sided test)		,178

Figure 8-2: Strategy Method versus Transparent Algorithm

Figure 8-2 shows the output of the independent-samples Kolmogorov-Smirnov test which compares the distribution of the amount sent in the strategy method condition and the transparent algorithm treatment.

8.2.1.3. Strategy Method versus Black-Box Algorithm



Total N		400
Most Extreme Differences	Absolute	,125
	Positive	,000
	Negative	-,125
Test Statistic		1,250
Asymptotic Sig. (2-sided test)		,088

Figure 8-3: Strategy Method versus Black-Box Algorithm

Figure 8-3 shows the output of the independent-samples Kolmogorov-Smirnov test which compares the distribution of the amount sent in the strategy method condition and the black-box algorithm treatment.

8.2.2. Test of Normality: Amount sent

The amount sent was not normally distributed in any condition as assessed by both the Kolmogorov-Smirnov's test, Shapiro-Wiik's test (all $ps < 0.001$) and a visual inspection of the Q-Q Plots below.

Table 8-1: Test of Normality amount sent

Table 8-1 shows the output of the normality tests conducted on each condition. The amount sent were not normally distributed in any of the four conditions (all $ps < 0.001$).

Condition		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Amount Sent	Strategy Method (SM)	0,141	200	0,000	0,853	200	0,000
	Game Method (GM)	0,157	200	0,000	0,837	200	0,000
	Transparent Algorithm (TA)	0,183	200	0,000	0,858	200	0,000
	Black-Box Algorithm (BBA)	0,16 9	200	0,000	0,870	200	0,000

a. Lilliefors Significance Correction

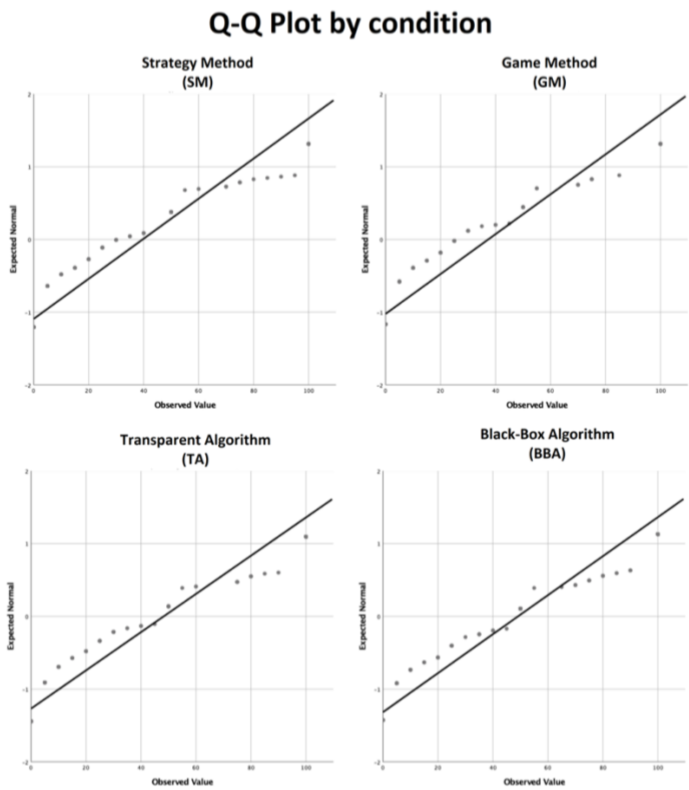


Figure 8-4: Q-Q plots amount sent by condition

Figure 8-4 illustrates a Q-Q plot of the amount sent in each condition. The amount sent is not normally distributed in any condition, as illustrated by the deviations from the diagonal line in each plot.

8.3. Appendix C: Instructions ‘Study 2’

8.3.1. Mturk HIT description

Economics Experiment for Research Purposes

Key information about the HIT:

- The HIT will take *approximately 15 minutes*.

Payment:

- You will be paid a $1.00 participation fee within three days upon completion of the HIT.
- You can earn a bonus that will be paid later.
- At the end of the experiment, you will receive a code to paste into the box below to receive your payment.

About the experiment:

- This HIT is an academic experiment on economic decision-making.
- Before the experiment starts you will have to read some instructions and answer some simple control questions.
- The experiment consists decision-making tasks and a questionnaire.

Browser compatibility:

- The experiment will not work using Internet Explorer.
- To perform the experiment, you must have Javascript and cookies enabled in your browser.

Voluntary participation and confidentiality:

- Participation in this HIT is entirely voluntary.
- Your answers are confidential and will not be revealed to anyone other than the experimenters.
- Your participation in the experiment is anonymous. No one, including the experimenters, will know any personal information about you.
- To know more about how we handle your privacy, please click here.

Informed consent:

- By accepting to participate in this HIT you give us informed consent that we can use your answers in an anonymized form for research purposes only.

Other information:

- Your Worker ID will be retrieved automatically when you click the link to the external survey page. Your Worker ID will only be used to secure payment to the right account and to control that you have not participated in this study before.
- **Make sure to leave this window open as you complete the experiment.** When you are finished, you must return to this page to paste the code into the box below and submit the HIT.

8.3.2. Welcome page – Stage 1

Welcome!

This HIT consists of two parts. The first part is an economics experiment where you will have to provide six price forecasts of financial assets. In the second part of the HIT, you have to answer a questionnaire containing 14 questions. You do not need to have any financial knowledge or background to complete the experiment.

You will be paid a participation fee of $1 for completing this HIT. In addition, you will be able to earn a bonus of $1 depending on the precision of your forecasts. At the last page of the HIT, you will receive a code to collect your payment via mTurk. The participation fee of $1 will be paid within three days upon completion, while the bonus will be paid soon after the whole experiment is conducted.

Before the experiment begins, you have to read some instructions. These instructions should be self-explanatory. However, if you have any questions or need clarifications, you should read back through the instructions. Once you have read the instructions, you will need to answer some simple control questions to make sure you have understood the task you are to perform.

During the HIT, please do not close this window or leave the HIT´s web pages in any other way. If you do close your browser or leave the HIT, you will not be able to re-enter, and we will not be able to pay you!

8.3.3. Instructions – Stage 2

Instructions

Please read the instructions below carefully and make sure you understand the experiment and the task you are to perform. In this experiment you will be asked to provide six price forecasts of the weekly closing price for different stocks traded on the New York Stock Exchange (NYSE). For each stock, you will be given a historical price graph for an undisclosed period that illustrates 52 weekly closing prices of the stock (one year). The first observation (t=1) represents the first week. The last observation (t=52) represents week 52. In addition, you will be given some financial information regarding the stock collected from week 52.

Your task is to forecast each stock's closing price four-weeks after the last observation (week 52). That is, you must forecast the stock's weekly closing price in week 56. We do not expect you to be familiar with valuation of financial assets to complete this task.

The bonus from the experiment is dependent on the precision of your forecasts. At the end of the experiment, one of the stocks you have provided your forecast for will be randomly chosen to determine your bonus. If your price forecast of the given stock is within 50 cents (either above or below) of the actual closing price in week 56, your bonus will be $1. If your forecast misses by more than 50 cents (either above or below) the actual closing price in week 56, you will not receive a bonus. If you have understood the instructions and are ready to proceed, please click on the button below. Note that you will not be able to return to this page.

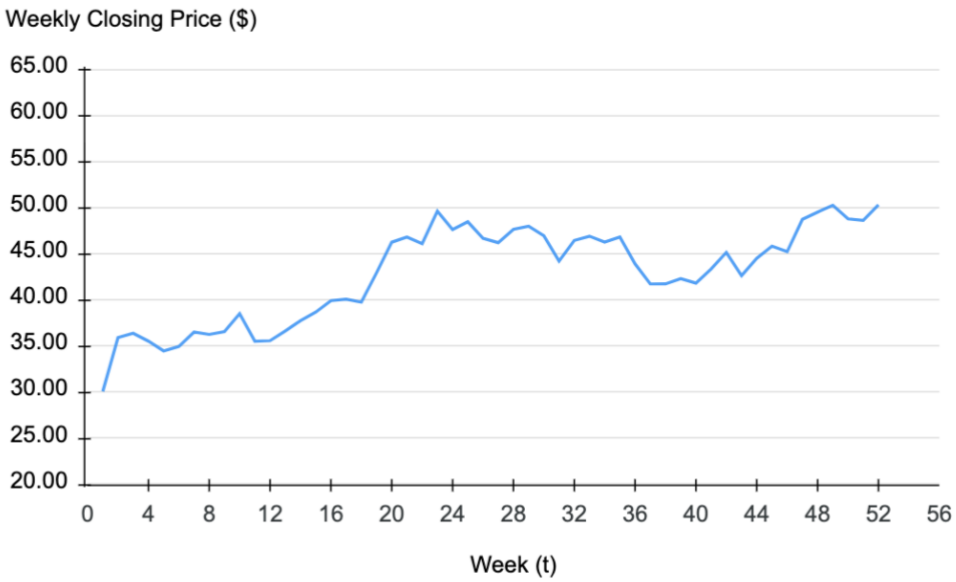
8.3.4. Pre-advice forecasting tasks – Stage 3, 4 and 5

8.3.4.1. Stock A

Stock A

Based on the information below, please provide your forecast of the weekly closing price for “Stock A” in week 56.

To display the weekly closing prices more easily, you can trace the graph with your cursor.



LAST FOUR OBSERVED WEEKLY CLOSING PRICES:

Week 49:	$50.27
Week 50:	$48.79
Week 51:	$48.63
Week 52:	$50.32

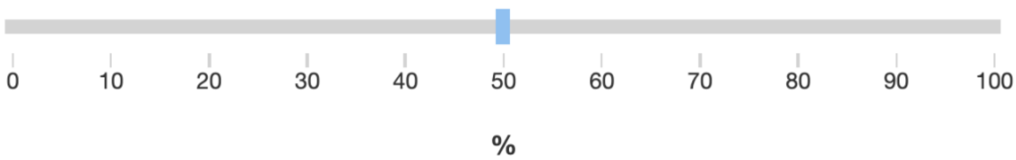
FINANCIAL INFORMATION:

Last 52-Week Low*:	$28.35
Last 52-Week High*:	$51.43
Dividend:	$0.10
P/E (TTM):	11.49
Beta (1 year):	2.094

* The 52-week low is the lowest daily closing price the stock had during the 52-week observation period.
* The 52-week high is the highest daily closing price the stock had during the 52-week observation period.

What is your forecast of the weekly closing price for “Stock A” in week 56?
Your answer may include two decimals.

How certain are you that your forecast is within 50 cents of the stock's actual weekly closing price in week 56?
Move the slider to indicate your level of confidence.
(0 means "Not certain at all", 100 means "Absolutely certain")

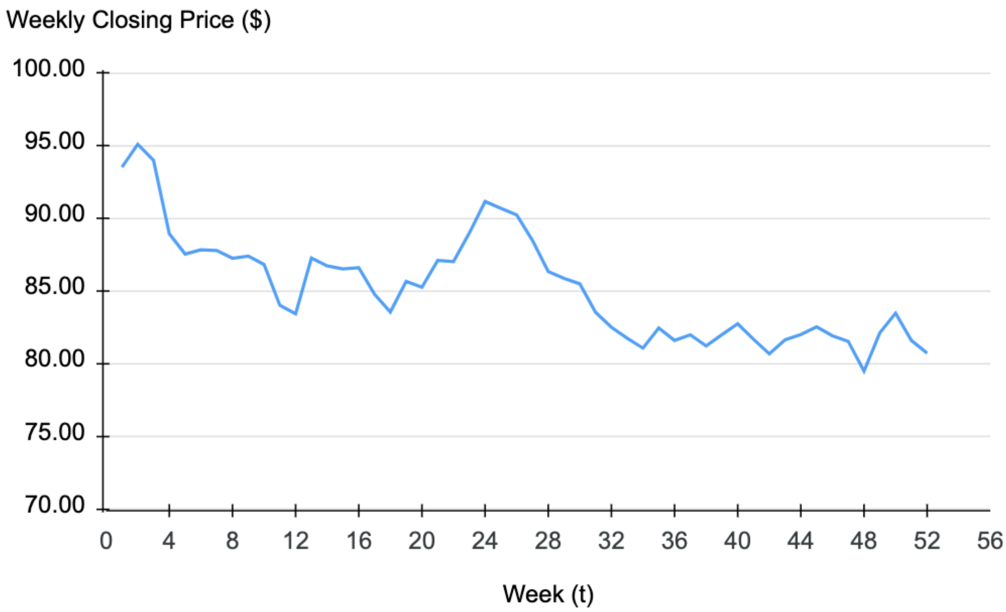


8.3.4.2. Stock B

Stock B

Based on the information below, please provide your forecast of the weekly closing price for "Stock B" in week 56.

To display the weekly closing prices more easily, you can trace the graph with your cursor.



LAST FOUR OBSERVED WEEKLY CLOSING PRICES:

Week 49:	$82.13
Week 50:	$83.49
Week 51:	$81.61
Week 52:	$80.73

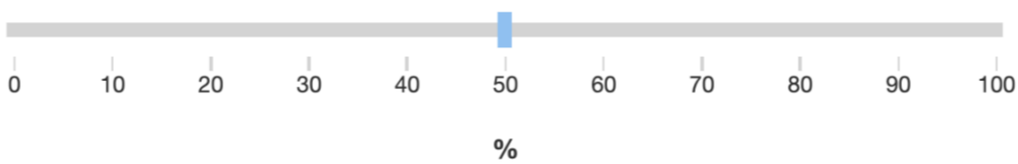
FINANCIAL INFORMATION:

Last 52-Week Low*:	$79.50
Last 52-Week High*:	$95.12
Dividend:	$0.77
P/E (TTM):	29.14
Beta (1 year):	0.2649

* The 52-week low is the lowest daily closing price the stock had during the 52-week observation period.
* The 52-week high is the highest daily closing price the stock had during the 52-week observation period.

What is your forecast of the weekly closing price for "Stock B" in week 56?
Your answer may include two decimals.

How certain are you that your forecast is within 50 cents of the stock's actual weekly closing price in week 56?
Move the slider to indicate your level of confidence.
(0 means "Not certain at all", 100 means "Absolutely certain")

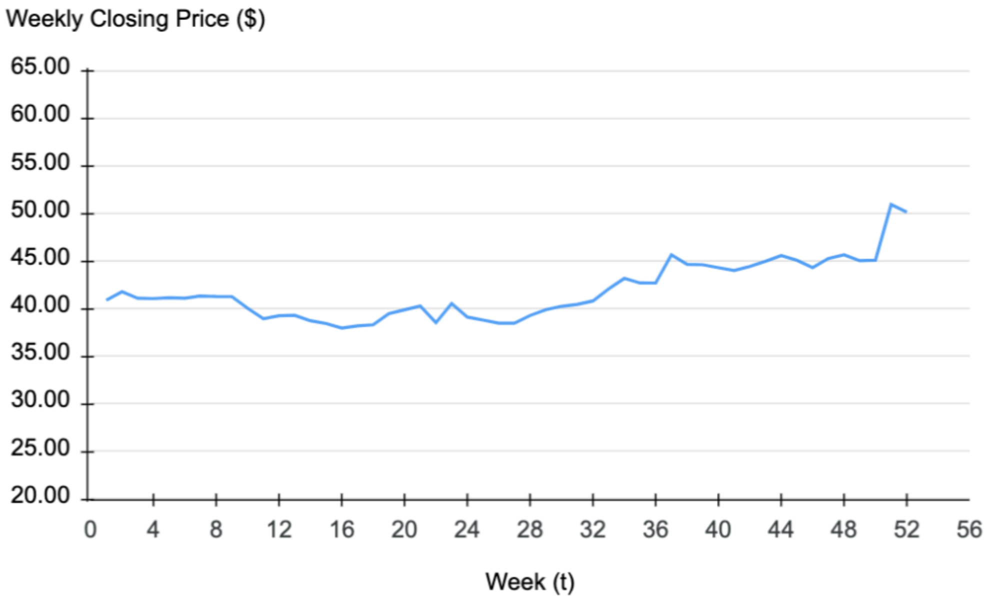


8.3.4.3. Stock C

Stock C

Based on the information below, please provide your forecast of the weekly closing price for “Stock C” in week 56.

To display the weekly closing prices more easily, you can trace the graph with your cursor.



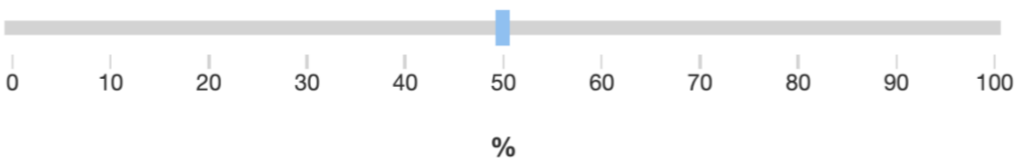
LAST FOUR OBSERVED WEEKLY CLOSING PRICES:	
Week 49:	$45.03
Week 50:	$45.09
Week 51:	$50.95
Week 52:	$50.14

FINANCIAL INFORMATION:	
Last 52-Week Low*:	$37.93
Last 52-Week High*:	$50.95
Dividend:	$0.19
P/E (TTM):	22.79
Beta (1 year):	0.7368

* The 52-week low is the lowest daily closing price the stock had during the 52-week observation period.
* The 52-week high is the highest daily closing price the stock had during the 52-week observation period.

What is your forecast of the weekly closing price for “Stock C” in week 56?
Your answer may include two decimals.

How certain are you that your forecast is within 50 cents of the stock's actual weekly closing price in week 56?
Move the slider to indicate your level of confidence.
(0 means "Not certain at all", 100 means "Absolutely certain")



8.3.5. Introduction of Advisor – Stage 6

8.3.5.1. Financial advisor condition

Additional Information

Before conducting this experiment, a **financial advisor** provided advice of the weekly closing price for stock A, B, and C for week 56. On the next few pages, these will be revealed to you. For each stock, you will be given the opportunity to adjust your initial forecast. The potential bonus from the experiment will solely depend on the precision of the final forecast you provide.

How useful do you think you will find the advice given by the financial advisor in helping you make your final estimate?

Please grade how useful you think you will find the advice ranging between 1 and 7 where 1 means "Not useful at all" and 7 means "Extremely useful".

1 2 3 4 5 6 7

8.3.5.2. Robo-advisor condition

Additional Information

Before conducting this experiment, a **financial robo-advisor** provided advice of the weekly closing price for stock A, B, and C for week 56. On the next few pages, these will be revealed to you. For each stock, you will be given the opportunity to adjust your initial forecast. The potential bonus from the experiment will solely depend on the precision of the final forecast you provide.

How useful do you think you will find the advice given by the financial robo-advisor in helping you make your final estimate?

Please grade how useful you think you will find the advice ranging between 1 and 7 where 1 means "Not useful at all" and 7 means "Extremely useful".

1 2 3 4 5 6 7

8.3.6. Post-advice forecasting tasks – Stage 7, 8 and 9

8.3.6.1. Stock A – Financial Advisor condition

Stock A

Based on the information you received about "Stock A", you forecasted the weekly closing price in week 56 to be **$ XX.XX**

Based on the same information, a **financial advisor** forecasted the weekly closing price in week 56 to be **$50.75**.

Click here to view the financial information

Your initial forecast: $ XX.XX

Advice from the financial advisor: $ 50.75

Please provide your final estimate of the weekly closing price for "Stock A" in week 56:
Your answer may include two decimals.

How certain are you that your forecast is within 50 cents of the stock's actual weekly closing price in week 56?
Move the slider to indicate your level of confidence.
(0 means "Not certain at all", 100 means "Absolutely certain")

0 10 20 30 40 50 60 70 80 90 100

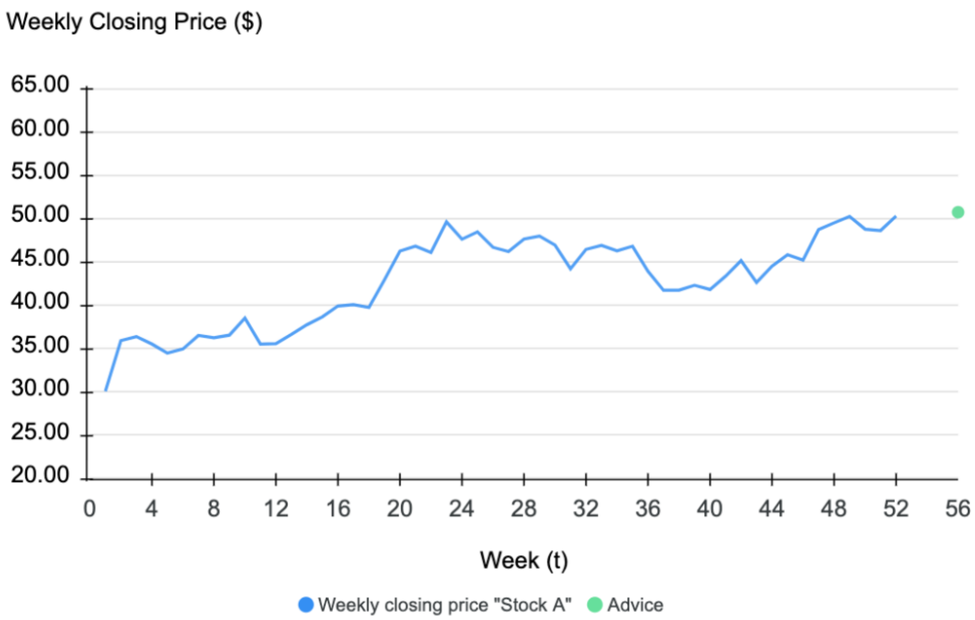
%

8.3.6.2. Stock A – Robo-advisor condition

Stock A

Based on the information you received about "Stock A", you forecasted the weekly closing price in week 56 to be **$ XX.XX**

Based on the same information, a **financial robo-advisor** forecasted the weekly closing price in week 56 to be **$50.75**.



Click here to view the financial information

Your initial forecast: $ XX.XX

Advice from the financial robo-advisor: $ 50.75

Please provide your final estimate of the weekly closing price for "Stock A" in week 56:
Your answer may include two decimals.

How certain are you that your forecast is within 50 cents of the stock's actual weekly closing price in week 56?
Move the slider to indicate your level of confidence.
(0 means "Not certain at all", 100 means "Absolutely certain")

0 10 20 30 40 50 60 70 80 90 100

%

8.3.6.3. Stock B – Financial Advisor condition

Stock B

Based on the information you received about "Stock B", you forecasted the weekly closing price in week 56 to be **$ XX.XX**

Based on the same information, a **financial advisor** forecasted the weekly closing price in week 56 to be **$80.35**.

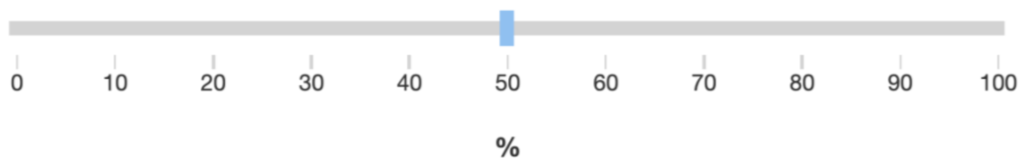
Click here to view the financial information

Your initial forecast: $ XX.XX

Advice from the financial advisor: $ 80.35

Please provide your final estimate of the weekly closing price for "Stock B" in week 56:
Your answer may include two decimals.

How certain are you that your forecast is within 50 cents of the stock's actual weekly closing price in week 56?
Move the slider to indicate your level of confidence.
(0 means "Not certain at all", 100 means "Absolutely certain")

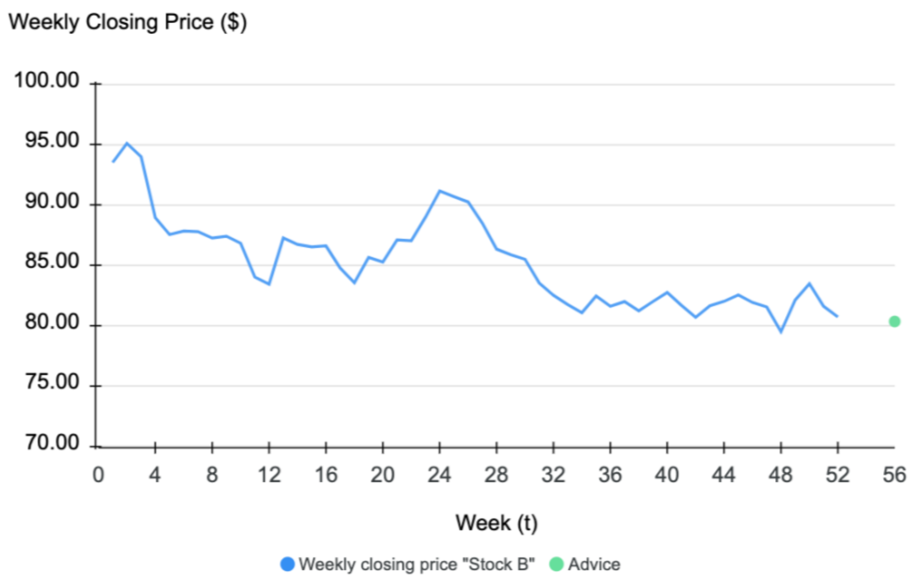


8.3.6.4. Stock B – Robo-advisor condition

Stock B

Based on the information you received about "Stock B", you forecasted the weekly closing price in week 56 to be **$XX.XX**

Based on the same information, a **financial robo-advisor** forecasted the weekly closing price in week 56 to be **$80.35**.



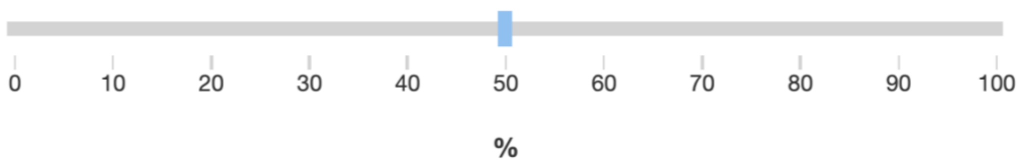
Click here to view the financial information

Your initial forecast: $ XX.XX

Advice from the financial robo-advisor: $ 80.35

Please provide your final estimate of the weekly closing price for "Stock B" in week 56:
Your answer may include two decimals.

How certain are you that your forecast is within 50 cents of the stock's actual weekly closing price in week 56?
Move the slider to indicate your level of confidence.
(0 means "Not certain at all", 100 means "Absolutely certain")

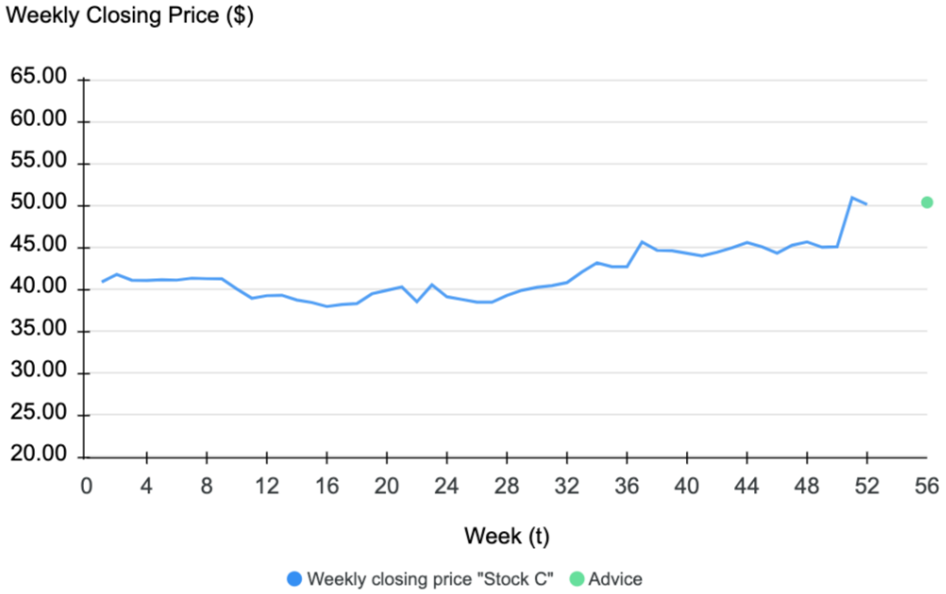


8.3.6.5. Stock C – Financial Advisor condition

Stock C

Based on the information you received about "Stock C", you forecasted the weekly closing price in week 56 to be **$ XX.XX**

Based on the same information, a **financial advisor** forecasted the weekly closing price in week 56 to be **$50.39**.



Click here to view the financial information

Your initial forecast:	$ XX.XX	Advice from the financial advisor:	$50.39

Please provide your final estimate of the weekly closing price for "Stock C" in week 56:
Your answer may include two decimals.

How certain are you that your forecast is within 50 cents of the stock's actual weekly closing price in week 56?
Move the slider to indicate your level of confidence.
(0 means "Not certain at all", 100 means "Absolutely certain")

0 10 20 30 40 50 60 70 80 90 100

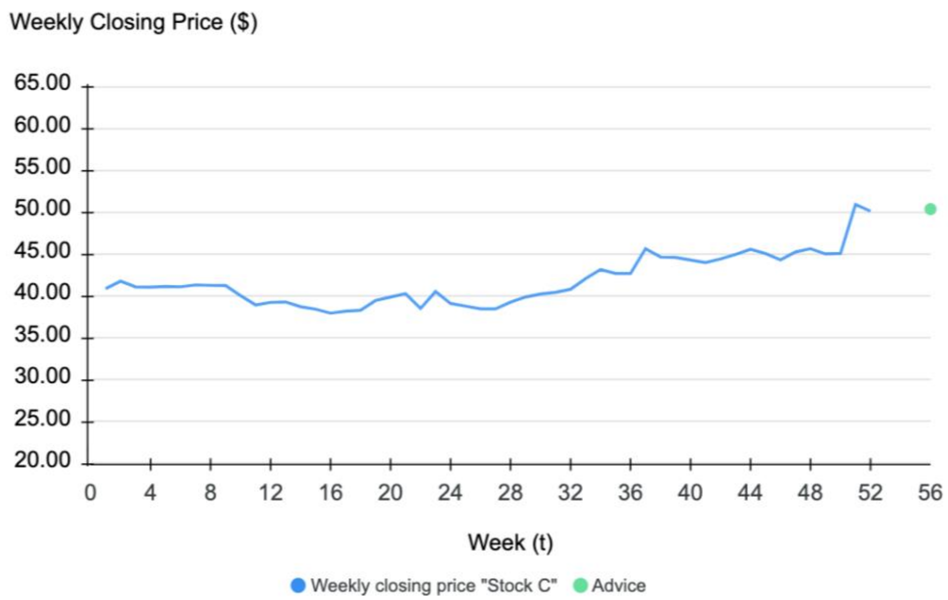
%

8.3.6.6. Stock C – Robo-advisor condition

Stock C

Based on the information you received about "Stock C", you forecasted the weekly closing price in week 56 to be **$ XX.XX**

Based on the same information, a **financial robo-advisor** forecasted the weekly closing price in week 56 to be **$50.39**.

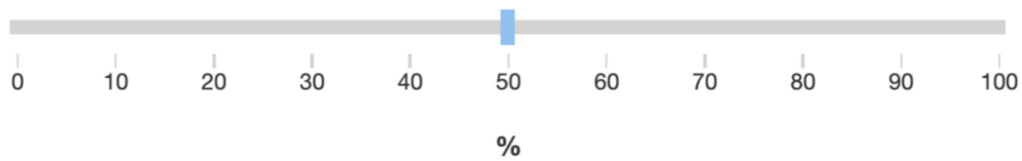


Click here to view the financial information

Your initial forecast: $ XX.XX Advice from the financial robo-advisor: $50.39

Please provide your final estimate of the weekly closing price for "Stock C" in week 56:
Your answer may include two decimals.

How certain are you that your forecast is within 50 cents of the stock's actual weekly closing price in week 56?
Move the slider to indicate your level of confidence.
(0 means "Not certain at all", 100 means "Absolutely certain")



8.3.7. Hyperlinks for the experiment `Study 2`

Financial advisor condition:

https://start.econexperiment.org/experiment/_beginParticipant.php?workerID=CG

Robo-advisor condition:

https://start.econexperiment.org/experiment/_beginParticipant.php?workerID=T1

8.4. Appendix D: Statistical tests ‘Study 2’

8.4.1. Output Hotelling’s T^2

Table 8-2: Output Hotelling’s T2

Dependent variables are SHIFT Stock A, SHIFT Stock B and SHIFT Stock C. The independent variable (condition) is the source of the advice, either a financial advisor or a robo-advisor respectively.

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	0,798	452,597[b]	3,000	343,000	0,000	0,798
	Wilks' Lambda	0,202	452,597[b]	3,000	343,000	0,000	0,798
	Hotelling's Trace	3,959	452,597[b]	3,000	343,000	0,000	0,798
	Roy's Largest Root	3,959	452,597[b]	3,000	343,000	0,000	0,798
Condition	Pillai's Trace	0,012	1,335[b]	3,000	343,000	0,263	0,012
	Wilks' Lambda	0,988	1,335[b]	3,000	343,000	0,263	0,012
	Hotelling's Trace	0,012	1,335[b]	3,000	343,000	0,263	0,012
	Roy's Largest Root	0,012	1,335[b]	3,000	343,000	0,263	0,012

a. Design: Intercept + Condition

b. Exact statistic

8.4.2. Assumptions Hotelling’s T^2

8.4.2.1. Linearity between the dependent variables for each group the independent variable

There was an approximately linear relationship between the SHIFT-scores for all stocks in each condition, as assessed by the scatterplots below.

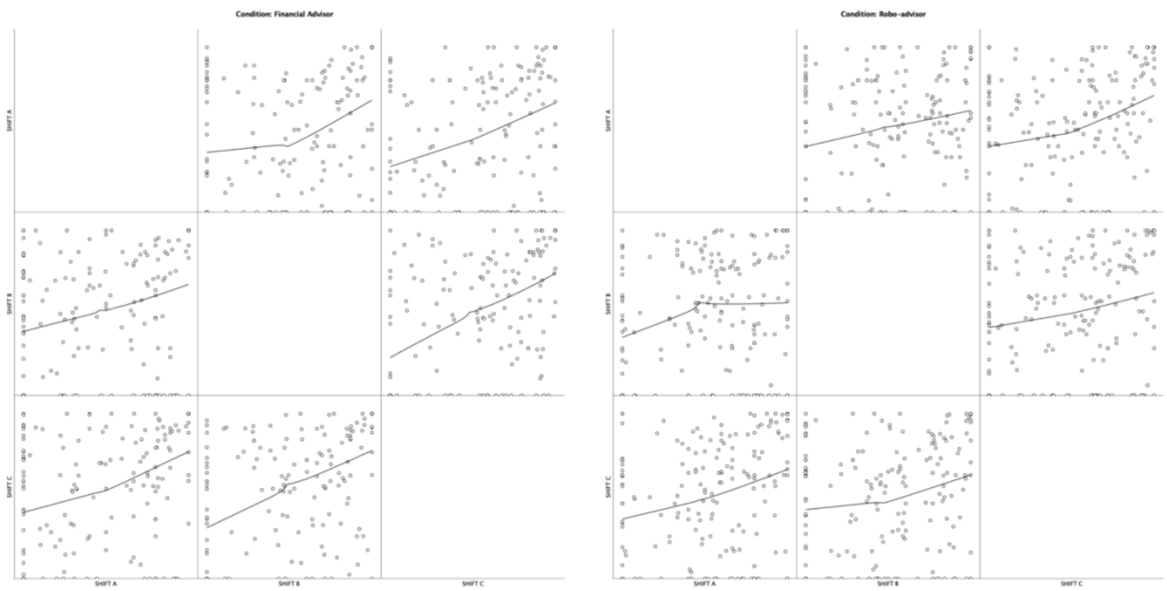


Figure 8-5: Scatterplot Matrix SHIFT-scores

Figure 8-5 illustrate a scatterplot matrix of the three SHIFT-scores. By looking at the lines in each scatterplot, we see that there was an approximately linear relationship between the SHIFT-scores for all stocks.

8.4.2.2. No multicollinearity

There was no evidence of multicollinearity (Pearson correlation, |r| < 0.9) (Tabachnick & Fidell, 2012).

Table 8-3: Pearson Correlation Matrix by condition

Table 8-3 shows the Pearson correlation between the SHIFT-scores in each condition. There was no evidence of multicollinearity as all the Person correlations were less then |0.90|.

Source = Financial Advisor

Correlations[a]

		SHIFT A	SHIFT B	SHIFT C
SHIFT A	Pearson Correlation	1	,260**	,317**
	Sig. (2-tailed)		0,001	0,000
	N	166	166	166
SHIFT B	Pearson Correlation	,260**	1	,450**
	Sig. (2-tailed)	0,001		0,000
	N	166	169	169
SHIFT C	Pearson Correlation	,317**	,450**	1
	Sig. (2-tailed)	0,000	0,000	
	N	166	169	169

**. Correlation is significant at the 0.01 level (2-tailed).

a. Source = Financial Advisor

Source = Robo-advisor

Correlations[a]

		SHIFT A	SHIFT B	SHIFT C
SHIFT A	Pearson Correlation	1	,215**	,271**
	Sig. (2-tailed)		0,004	0,000
	N	181	181	181
SHIFT B	Pearson Correlation	,215**	1	,178*
	Sig. (2-tailed)	0,004		0,017
	N	181	181	181
SHIFT C	Pearson Correlation	,271**	,178*	1
	Sig. (2-tailed)	0,000	0,017	
	N	181	181	181

**. Correlation is significant at the 0.01 level (2-tailed).

a. Source = Robo-advisor

8.4.2.3. No univariate outliers

There were no univariate outliers in the data, as assessed by inspection of a boxplot.

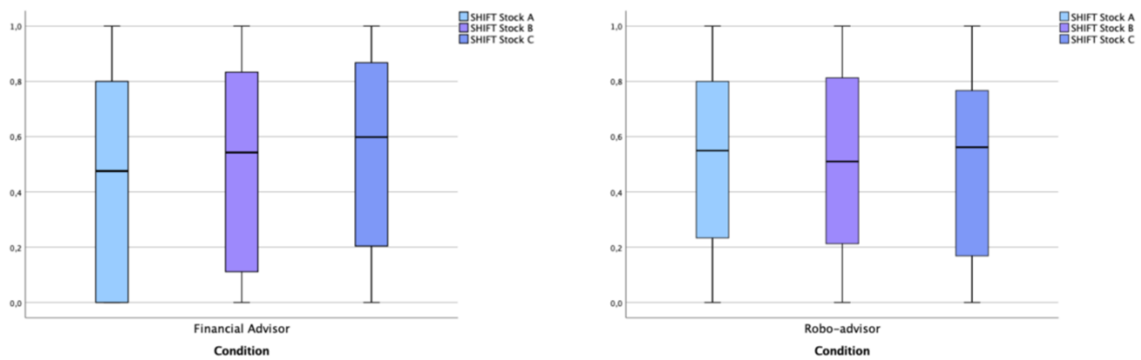


Figure 8-6: Boxplots by condition

Figure 8-6 shows a boxplot of the SHIFT-scores for each condition. By assessing the boxplots, we can see that there were no univariate outliers in the data.

8.4.2.4. No multivariate outliers

The maximum observed Mahalanobis distance was 9.108. The critical value for Mahalanobis distance with three dependent variables is 16.267. Hence, there were no multivariate outliers in the data (all $ps > 0.0278$).

8.4.2.5. Multivariate normality - Violated

The SHIFT-scores was not normally distributed in any condition as assessed by both Kolmogorov-Smirnov's test $(p < 0.001)$ and a visual inspection of the Q-Q Plots below.

Table 8-4: Test of normality SHIFT by condition

Table 8-4 shows the output of the normality tests for each condition. None of the SHIFT-scores were normality distributed in any of the conditions (all $ps < 0.001$).

Tests of Normality[a]

Condition		Kolmogorov-Smirnov[b] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
SHIFT Stock A	Financial Advisor	0,152	166	0,000	0,882	166	0,000
SHIFT Stock B	Financial Advisor	0,159	166	0,000	0,889	166	0,000
SHIFT Stock C	Financial Advisor	0,112	166	0,000	0,895	166	0,000

a. Condition = Financial Advisor

b. Lilliefors Significance Correction

Tests of Normality[a]

Condition		Kolmogorov-Smirnov[b] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
SHIFT Stock A	Robo-advisor	0,135	181	0,000	0,910	181	0,000
SHIFT Stock B	Robo-advisor	0,141	181	0,000	0,907	181	0,000
SHIFT Stock C	Robo-advisor	0,120	181	0,000	0,915	181	0,000

a. Condition = Robo-advisor

b. Lilliefors Significance Correction

Q-Q plots Financial Advisor

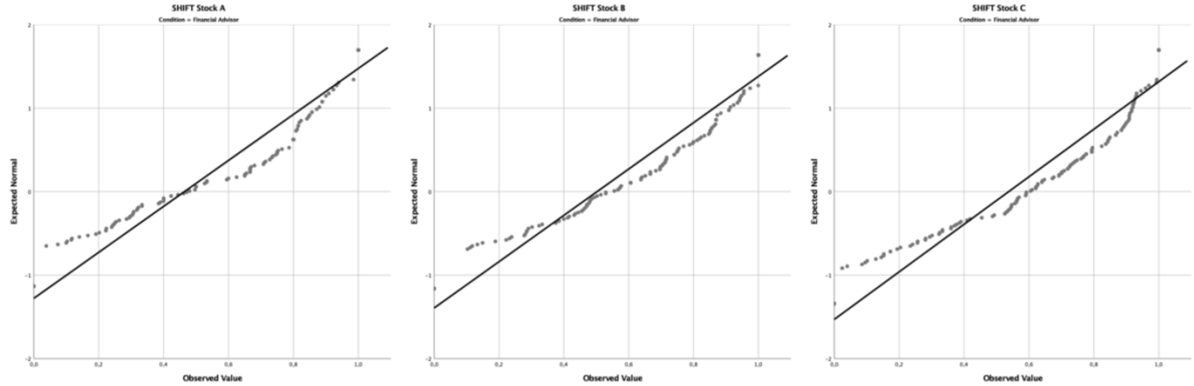


Figure 8-7: Q-Q plots, financial advisor condition

Figure 8-7 illustrates a Q-Q plot of each SHIFT-score in the financial advisor condition. None of the SHIFT-scores were normally distributed, as illustrated by the deviations from the diagonal line in each plot.

Q-Q plots Robo-advisor

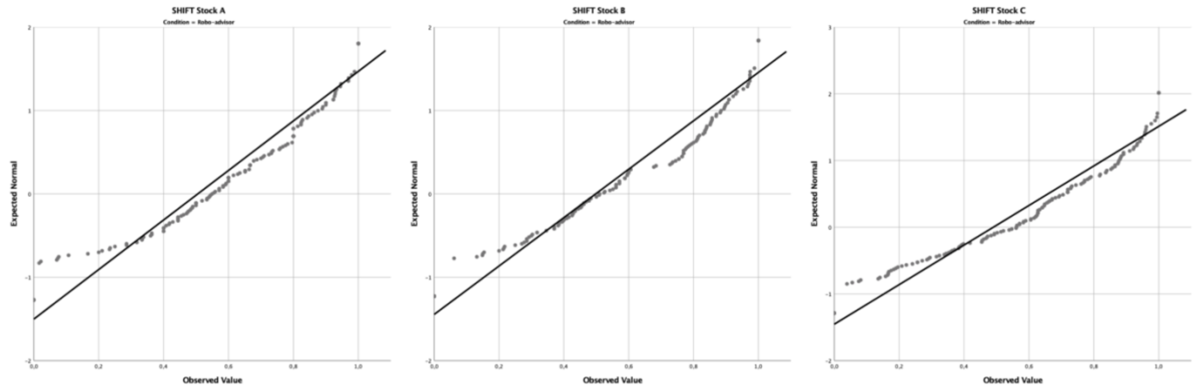


Figure 8-8: Q-Q plots, robo-advisor condition

Figure 8-8 illustrates a Q-Q plot of each SHIFT-score in the robo-advisor condition. None of the SHIFT-scores were normally distributed, as illustrated by the deviations from the diagonal line in each plot.

8.4.2.6. Homogeneity of variance-covariance

By looking at the table below, we can see that the assumption of homogeneity of variance-covariance matrices were met ($p = 0.178$).

Table 8-5: Box's M Test

Table 8-5 shows the output for the Box's test of equality of covariance matrices. By looking at the p-value, we find no significant difference in the covariance matrices in the two conditions.

Box's Test of Equality of Covariance Matrices[a]

Box's M	9,002
F	1,486
df1	6
df2	845193,004
Sig.	0,178

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + Source

8.4.2.7. Homogeneity of variances – Violated

By analyzing the Levene's Test of Homogeneity of Variance below, we can see that the SHIFT-score for Stock A has heterogenous variances. Thus, the homogeneity of variances assumption was violated.

Table 8-6: Levene´s Test

Table 8-6 shows the output for the Levene´s test of equality of error variances. By looking at the p-values, we can see that the SHIFT-score for Stock A has heterogeneous variances and therefore violates the assumption of homogeneity of variances.

Levene's Test of Equality of Error Variances[a]

		Levene Statistic	df1	df2	Sig.
SHIFT A	Based on Mean	5,855	1	345	0,016
	Based on Median	6,067	1	345	0,014
	Based on Median and with adjusted df	6,067	1	333,101	0,014
	Based on trimmed mean	5,884	1	345	0,016
SHIFT B	Based on Mean	1,017	1	345	0,314
	Based on Median	0,922	1	345	0,338
	Based on Median and with adjusted df	0,922	1	344,816	0,338
	Based on trimmed mean	1,015	1	345	0,314
SHIFT C	Based on Mean	0,441	1	345	0,507
	Based on Median	0,266	1	345	0,606
	Based on Median and with adjusted df	0,266	1	344,697	0,606
	Based on trimmed mean	0,382	1	345	0,537

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Condition

8.4.3. Output one-way MANCOVA

Table 8-7: Output one-way MANCOVA

Dependent variables are SHIFT Stock A, SHIFT Stock B and SHIFT Stock C. The independent variable (condition) is the source of the advice, either a financial advisor or a robo-advisor respectively.

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	0,118	14,526[b]	3,000	325,00	0,000	0,118
	Wilks' Lambda	0,882	14,526[b]	3,000	325,00	0,000	0,118
	Hotelling's Trace	0,134	14,526[b]	3,000	325,00	0,000	0,118
	Roy's Largest Root	0,134	14,526[b]	3,000	325,00	0,000	0,118
Condition	Pillai's Trace	0,013	1,388[b]	3,000	325,00	0,246	0,013
	Wilks' Lambda	0,987	1,388[b]	3,000	325,00	0,246	0,013
	Hotelling's Trace	0,013	1,388[b]	3,000	325,00	0,246	0,013
	Roy's Largest Root	0,013	1,388[b]	3,000	325,00	0,246	0,013
Gender	Pillai's Trace	0,013	0,729	6,000	652,00	0,626	0,007
	Wilks' Lambda	0,987	,729[b]	6,000	650,00	0,626	0,007
	Hotelling's Trace	0,013	0,729	6,000	648,00	0,627	0,007
	Roy's Largest Root	0,013	1,384[c]	3,000	326,00	0,248	0,013
Education level	Pillai's Trace	0,027	0,756	12,000	981,00	0,696	0,009
	Wilks' Lambda	0,973	0,753	12,000	860,16	0,700	0,009
	Hotelling's Trace	0,028	0,749	12,000	971,00	0,703	0,009
	Roy's Largest Root	0,014	1,152[c]	4,000	327,00	0,332	0,014
Comprehension questions	Pillai's Trace	0,029	3,265[b]	3,000	325,00	0,022	0,029
	Wilks' Lambda	0,971	3,265[b]	3,000	325,00	0,022	0,029
	Hotelling's Trace	0,030	3,265[b]	3,000	325,00	0,022	0,029
	Roy's Largest Root	0,030	3,265[b]	3,000	325,00	0,022	0,029
Perceived difficulty	Pillai's Trace	0,040	0,735	18,000	981,00	0,777	0,013
	Wilks' Lambda	0,960	0,734	18,000	919,72	0,778	0,013
	Hotelling's Trace	0,041	0,733	18,000	971,00	0,779	0,013
	Roy's Largest Root	0,028	1,520[c]	6,000	327,00	0,171	0,027
Overshooter	Pillai's Trace	0,026	2,939[b]	3,000	325,00	0,033	0,026
	Wilks' Lambda	0,974	2,939[b]	3,000	325,00	0,033	0,026
	Hotelling's Trace	0,027	2,939[b]	3,000	325,00	0,033	0,026
	Roy's Largest Root	0,027	2,939[b]	3,000	325,00	0,033	0,026
Age	Pillai's Trace	0,016	1,708[b]	3,000	325,00	0,165	0,016
	Wilks' Lambda	0,984	1,708[b]	3,000	325,00	0,165	0,016
	Hotelling's Trace	0,016	1,708[b]	3,000	325,00	0,165	0,016
	Roy's Largest Root	0,016	1,708[b]	3,000	325,00	0,165	0,016
Pre-Confidence Stock A	Pillai's Trace	0,016	1,706[b]	3,000	325,00	0,166	0,016
	Wilks' Lambda	0,984	1,706[b]	3,000	325,00	0,166	0,016
	Hotelling's Trace	0,016	1,706[b]	3,000	325,00	0,166	0,016
	Roy's Largest Root	0,016	1,706[b]	3,000	325,00	0,166	0,016
Pre-Confidence Stock B	Pillai's Trace	0,009	1,008[b]	3,000	325,00	0,389	0,009
	Wilks' Lambda	0,991	1,008[b]	3,000	325,00	0,389	0,009
	Hotelling's Trace	0,009	1,008[b]	3,000	325,00	0,389	0,009
	Roy's Largest Root	0,009	1,008[b]	3,000	325,00	0,389	0,009
Pre-Confidence Stock C	Pillai's Trace	0,020	2,182[b]	3,000	325,00	0,090	0,020
	Wilks' Lambda	0,980	2,182[b]	3,000	325,00	0,090	0,020
	Hotelling's Trace	0,020	2,182[b]	3,000	325,00	0,090	0,020
	Roy's Largest Root	0,020	2,182[b]	3,000	325,00	0,090	0,020

a. Design: Intercept + Condition + Gender + Education_Level + QC + Percieved_Difficulty + Overshooter + Age + Pre_Conf_A + Pre_Conf_B + Pre_Conf_C

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

8.4.4. Assumptions one-way MANCOVA

8.4.4.1. Linear relationship between the dependent variables in each condition

There was an approximately linear relationship between the SHIFT-scores for all stocks in each condition, as assessed by the scatterplots below.

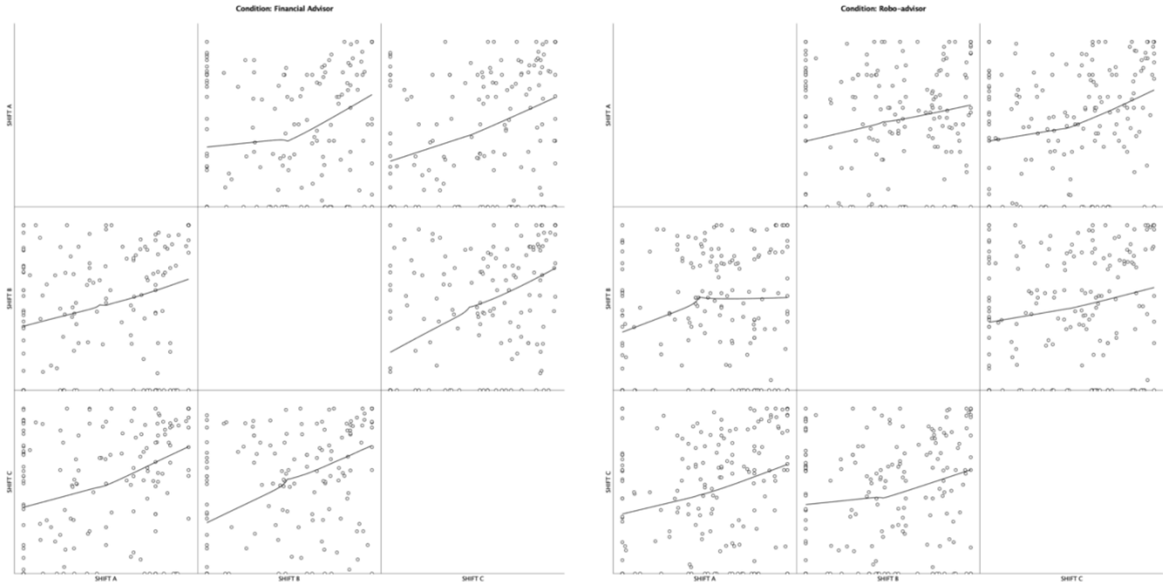


Figure 8-9: Scatterplot Matrix SHIFT-scores

Figure 8-9 illustrates a scatterplot matrix of the three SHIFT-scores for each condition. By looking at the lines in each scatterplot, we see that there was an approximately linear relationship between the SHIFT-scores for all stocks in each condition.

8.4.4.2. Linear relationship between the covariates and the dependent variables in each condition – Violated

There was not a linear relationship between the covariates and the dependent variables, as assessed by a visual inspection of the scatterplots below

Covariate: Age

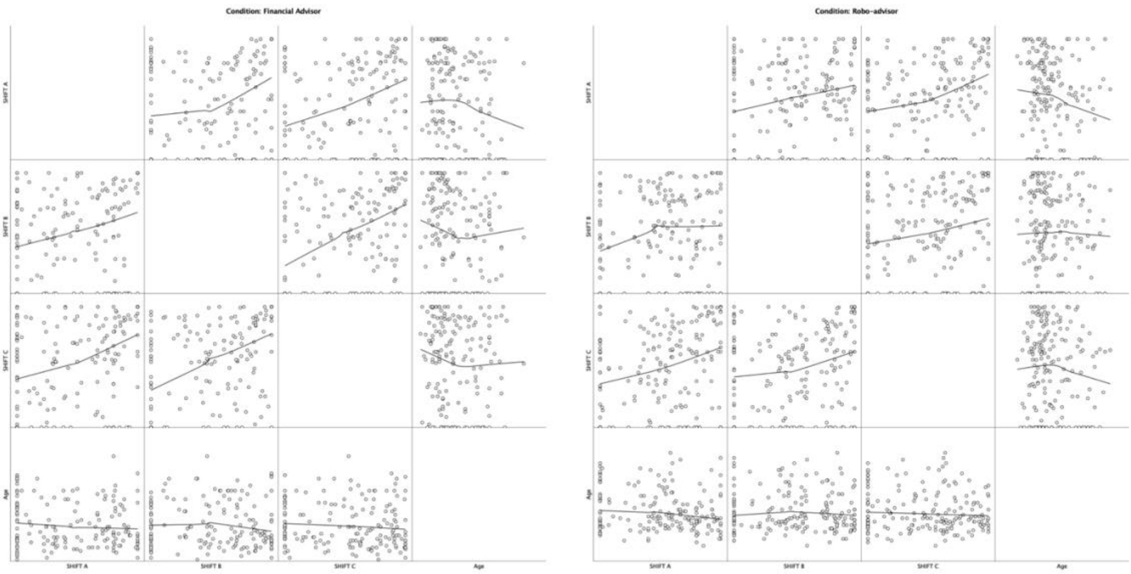


Figure 8-10: Scatterplot Matrix SHIFT-scores and age

Figure 8-10 illustrates a scatterplot matrix of the three SHIFT-scores and a covariate for age for each condition. By looking at the lines in each scatterplot, we see that there was no linear relationship between the SHIFT-scores and age for all stocks in each condition.

Covariate: Pre-Confidence Stock A

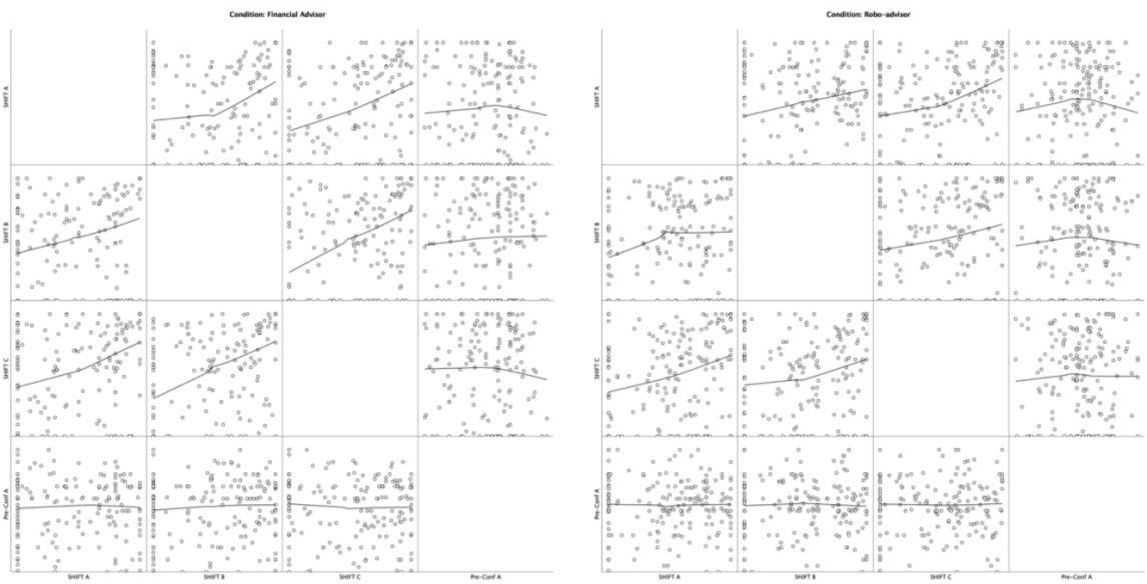


Figure 8-11: Scatterplot Matrix SHIFT-scores and pre-conf Stock A

Figure 8-11 illustrates a scatterplot matrix of the three SHIFT-scores and a covariate for age for each condition. By looking at the lines in each scatterplot, we see that there was no linear relationship between the SHIFT-scores and pre-confidence level of stock A for all stocks in each condition.

Covariate: Pre-Confidence Stock B

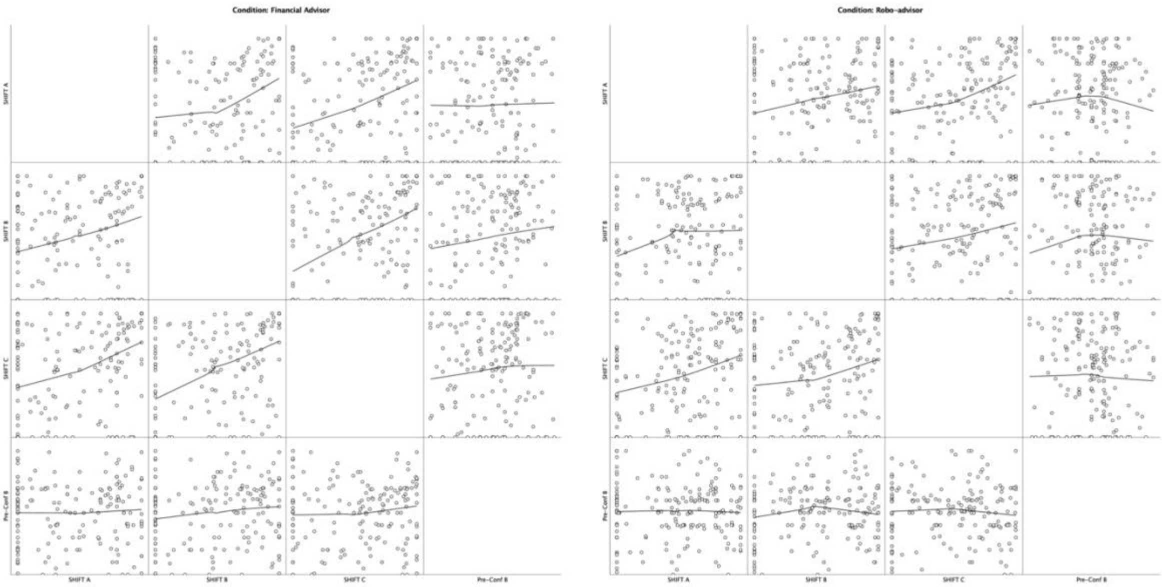


Figure 8-12: Scatterplot Matrix SHIFT-scores and pre-conf Stock B

Figure 8-12 illustrates a scatterplot matrix of the three SHIFT-scores and a covariate for age for each condition. By looking at the lines in each scatterplot, we see that there was no linear relationship between the SHIFT-scores and pre-confidence level of stock B for all stocks in each condition.

Covariate: Pre-Confidence Stock C

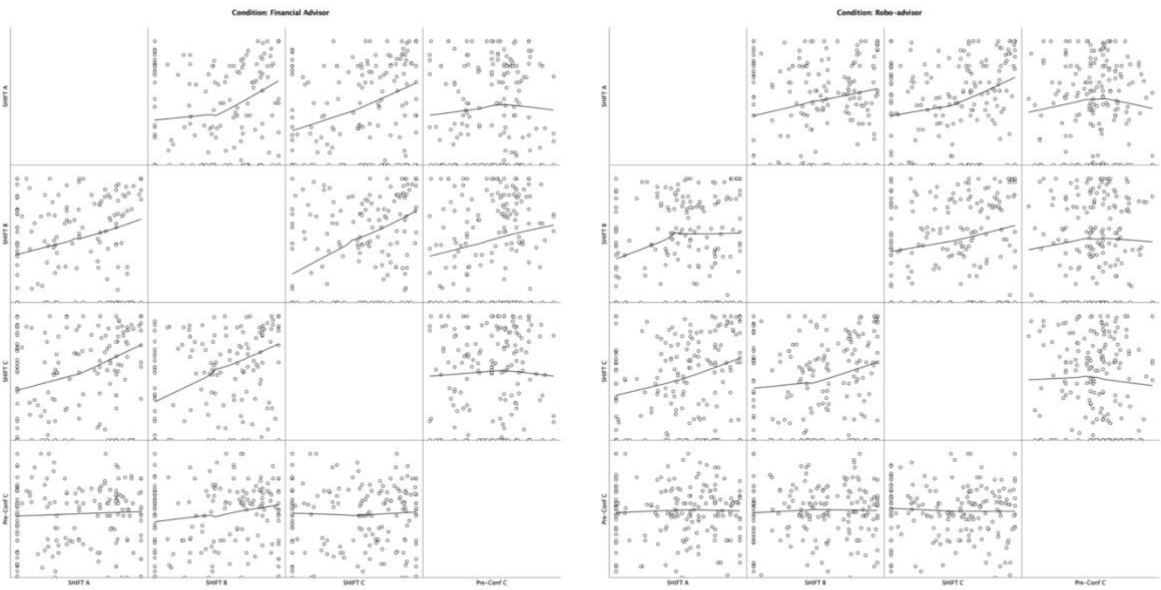


Figure 8-13: Scatterplot Matrix SHIFT-scores and pre-conf Stock C

Figure 8-13 illustrates a scatterplot matrix of the three SHIFT-scores and a covariate for age for each condition. By looking at the lines in each scatterplot, we see that there was no linear relationship between the SHIFT-scores and pre-confidence level of stock C for all stocks in each condition.

8.4.4.3. Homogeneity of regression slopes

Table 8-8: Test of homogeneity of regression slopes

There was heterogeneity of regression slopes, as assessed by the interaction terms between Age and Condition, $F(3, 321) = 0.102, p = 0.959$, Pre_Conf_A and Condition, $F(3, 321) = 1.222, p = 0.302$, Pre_Conf_B and Condition, $F(3, 321) = 1.753, p = 0.156$ and Pre_Conf_C and Condition $F(3, 321) = 0.717, p = 0.543$.

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	0,118	14,249[b]	3,000	321,000	0,000	0,118
	Wilks' Lambda	0,882	14,249[b]	3,000	321,000	0,000	0,118
	Hotelling's Trace	0,133	14,249[b]	3,000	321,000	0,000	0,118
	Roy's Largest Root	0,133	14,249[b]	3,000	321,000	0,000	0,118
Condition	Pillai's Trace	0,001	,088[b]	3,000	321,000	0,966	0,001
	Wilks' Lambda	0,999	,088[b]	3,000	321,000	0,966	0,001
	Hotelling's Trace	0,001	,088[b]	3,000	321,000	0,966	0,001
	Roy's Largest Root	0,001	,088[b]	3,000	321,000	0,966	0,001
Gender	Pillai's Trace	0,013	0,685	6,000	644,000	0,662	0,006
	Wilks' Lambda	0,987	,685[b]	6,000	642,000	0,662	0,006
	Hotelling's Trace	0,013	0,684	6,000	640,000	0,662	0,006
	Roy's Largest Root	0,012	1,300[c]	3,000	322,000	0,274	0,012
Education Level	Pillai's Trace	0,026	0,701	12,000	969,000	0,752	0,009
	Wilks' Lambda	0,974	0,697	12,000	849,578	0,755	0,009
	Hotelling's Trace	0,026	0,694	12,000	959,000	0,758	0,009
	Roy's Largest Root	0,014	1,090[c]	4,000	323,000	0,361	0,013
Comprehension Questions	Pillai's Trace	0,029	3,189[b]	3,000	321,000	0,024	0,029
	Wilks' Lambda	0,971	3,189[b]	3,000	321,000	0,024	0,029
	Hotelling's Trace	0,030	3,189[b]	3,000	321,000	0,024	0,029
	Roy's Largest Root	0,030	3,189[b]	3,000	321,000	0,024	0,029
Percieved Difficulty	Pillai's Trace	0,041	0,755	18,000	969,000	0,755	0,014
	Wilks' Lambda	0,959	0,754	18,000	908,410	0,756	0,014
	Hotelling's Trace	0,042	0,753	18,000	959,000	0,757	0,014
	Roy's Largest Root	0,028	1,498[c]	6,000	323,000	0,178	0,027
Overshooter	Pillai's Trace	0,025	2,692[b]	3,000	321,000	0,046	0,025
	Wilks' Lambda	0,975	2,692[b]	3,000	321,000	0,046	0,025
	Hotelling's Trace	0,025	2,692[b]	3,000	321,000	0,046	0,025
	Roy's Largest Root	0,025	2,692[b]	3,000	321,000	0,046	0,025
Age	Pillai's Trace	0,014	1,563[b]	3,000	321,000	0,198	0,014
	Wilks' Lambda	0,986	1,563[b]	3,000	321,000	0,198	0,014
	Hotelling's Trace	0,015	1,563[b]	3,000	321,000	0,198	0,014
	Roy's Largest Root	0,015	1,563[b]	3,000	321,000	0,198	0,014
Pre_Conf_A	Pillai's Trace	0,019	2,064[b]	3,000	321,000	0,105	0,019
	Wilks' Lambda	0,981	2,064[b]	3,000	321,000	0,105	0,019
	Hotelling's Trace	0,019	2,064[b]	3,000	321,000	0,105	0,019
	Roy's Largest Root	0,019	2,064[b]	3,000	321,000	0,105	0,019
Pre_Conf_B	Pillai's Trace	0,007	,785[b]	3,000	321,000	0,503	0,007
	Wilks' Lambda	0,993	,785[b]	3,000	321,000	0,503	0,007
	Hotelling's Trace	0,007	,785[b]	3,000	321,000	0,503	0,007
	Roy's Largest Root	0,007	,785[b]	3,000	321,000	0,503	0,007
Pre_Conf_C	Pillai's Trace	0,025	2,693[b]	3,000	321,000	0,046	0,025
	Wilks' Lambda	0,975	2,693[b]	3,000	321,000	0,046	0,025
	Hotelling's Trace	0,025	2,693[b]	3,000	321,000	0,046	0,025
	Roy's Largest Root	0,025	2,693[b]	3,000	321,000	0,046	0,025
Condition * Age	Pillai's Trace	0,001	,102[b]	3,000	321,000	0,959	0,001
	Wilks' Lambda	0,999	,102[b]	3,000	321,000	0,959	0,001
	Hotelling's Trace	0,001	,102[b]	3,000	321,000	0,959	0,001
	Roy's Largest Root	0,001	,102[b]	3,000	321,000	0,959	0,001
Condition * Pre_Conf_A	Pillai's Trace	0,011	1,222[b]	3,000	321,000	0,302	0,011
	Wilks' Lambda	0,989	1,222[b]	3,000	321,000	0,302	0,011
	Hotelling's Trace	0,011	1,222[b]	3,000	321,000	0,302	0,011
	Roy's Largest Root	0,011	1,222[b]	3,000	321,000	0,302	0,011
Condition * Pre_Conf_B	Pillai's Trace	0,016	1,753[b]	3,000	321,000	0,156	0,016
	Wilks' Lambda	0,984	1,753[b]	3,000	321,000	0,156	0,016
	Hotelling's Trace	0,016	1,753[b]	3,000	321,000	0,156	0,016
	Roy's Largest Root	0,016	1,753[b]	3,000	321,000	0,156	0,016
Condition * Pre_Conf_C	Pillai's Trace	0,007	,717[b]	3,000	321,000	0,543	0,007
	Wilks' Lambda	0,993	,717[b]	3,000	321,000	0,543	0,007
	Hotelling's Trace	0,007	,717[b]	3,000	321,000	0,543	0,007
	Roy's Largest Root	0,007	,717[b]	3,000	321,000	0,543	0,007

a. Design: Intercept + AdviceSource + Gender + Education_Level + QC + Percieved_Difficulty + Overshooter + Age +

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

8.4.4.4. No multicollinearity

There was no evidence of multicollinearity (Pearson correlation, $|r| < 0.9$) (Tabachnick & Fidell, 2012).

Table 8-9: Pearson Correlation Matrix

Table 8-9 shows the Pearson correlation between the SHIFT-scores in each condition. There was no evidence of multicollinearity as all the Person correlations were less then |0.90|.

Source = Financial Advisor

Correlations[a]

		SHIFT A	SHIFT B	SHIFT C
SHIFT A	Pearson Correlation	1	,260**	,317**
	Sig. (2-tailed)		0,001	0,000
	N	166	166	166
SHIFT B	Pearson Correlation	,260**	1	,450**
	Sig. (2-tailed)	0,001		0,000
	N	166	169	169
SHIFT C	Pearson Correlation	,317**	,450**	1
	Sig. (2-tailed)	0,000	0,000	
	N	166	169	169

**. Correlation is significant at the 0.01 level (2-tailed).

a. Source = Financial Advisor

Source = Robo-advisor

Correlations[a]

		SHIFT A	SHIFT B	SHIFT C
SHIFT A	Pearson Correlation	1	,215**	,271**
	Sig. (2-tailed)		0,004	0,000
	N	181	181	181
SHIFT B	Pearson Correlation	,215**	1	,178*
	Sig. (2-tailed)	0,004		0,017
	N	181	181	181
SHIFT C	Pearson Correlation	,271**	,178*	1
	Sig. (2-tailed)	0,000	0,017	
	N	181	181	181

**. Correlation is significant at the 0.01 level (2-tailed).

a. Source = Robo-advisor

8.4.4.5. No univariate outliers

There were no univariate outliers in the data. This was examined by analyzing the standardized residuals. No standardized residuals were greater than +/- 3 standard deviations.

8.4.4.6. No multivariate outliers

The maximum observed Mahalanobis distance was 9.108. The critical value for Mahalanobis distance with three dependent variables is 16.267. Hence, there were no multivariate outliers in the data (all $ps > 0.0278$).

8.4.4.7. Multivariate normality - Violated

The residuals for the SHIFT-scores was not normally distributed in any condition as assessed by both Kolmogorov-Smirnov's test ($p < 0.001$) and a visual inspection of the Q-Q Plots below.

Table 8-10: Test of Normality

Table 8-10 shows the output of the normality tests for each condition. None of the SHIFT-scores were normality distributed in any of the conditions (all $ps < 0.01$).

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
Condition		Statistic	df	Sig.	Statistic	df	Sig.
Residual for SHIFT_A	Financial Advisor	0,103	166	0,000	0,936	166	0,000
	Robo-advisor	0,095	181	0,000	0,965	181	0,000
Residual for SHIFT_B	Financial Advisor	0,106	166	0,000	0,942	166	0,000
	Robo-advisor	0,105	181	0,000	0,951	181	0,000
Residual for SHIFT_C	Financial Advisor	0,102	166	0,000	0,921	166	0,000
	Robo-advisor	0,092	181	0,001	0,945	181	0,000

a. Lilliefors Significance Correction

Q.Q plots financial advisor condition

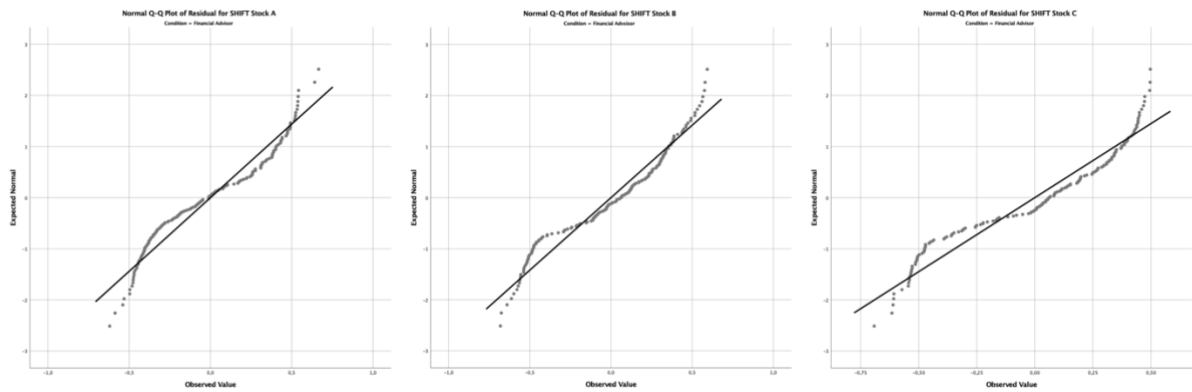


Figure 8-14: Q-Q-plots financial advisor condition

Figure 8-14 illustrates a Q-Q plot of each SHIFT-score in the financial advisor condition. None of the SHIFT-scores were normally distributed, as illustrated by the deviations from the diagonal line in each plot.

Q-Q plots robo-advisor condition

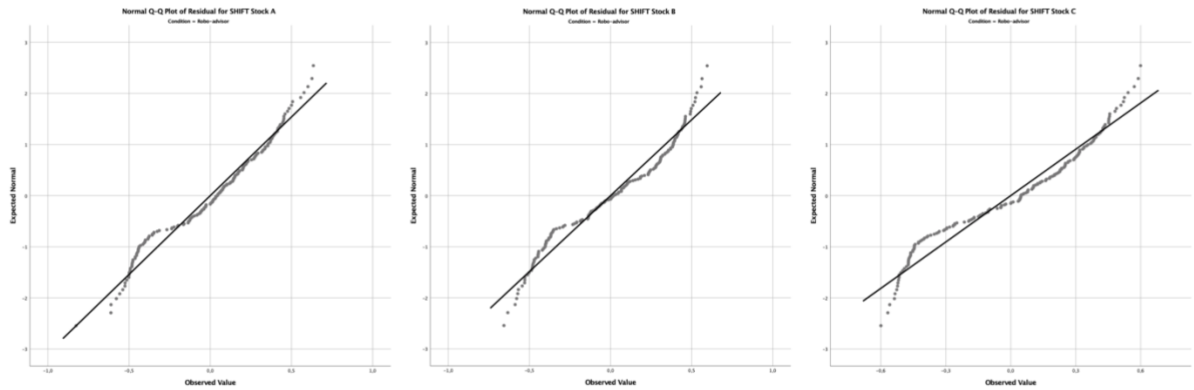


Figure 8-15: Q-Q-plots robo-advisor condition

Figure 8-15 illustrates a Q-Q plot of each SHIFT-score in the robo-advisor condition. None of the SHIFT-scores were normally distributed, as illustrated by the deviations from the diagonal line in each plot.

8.4.4.8. Homogeneity of variance-covariance

By looking at the table below, we can see that the assumption of homogeneity of variance-covariance matrices were met ($p = 0.796$).

Table 8-11: Box's M Test

Table 8-11 shows the output for the Box's test of equality of covariance matrices. By looking at the p-value, we find no significant difference in the covariance matrices in the two conditions.

Box's Test of Equality of Covariance Matrices[a]

Box's M	238,120
F	0,909
df1	174
df2	4264,621
Sig.	0,796

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + AdviceSource + Gender + Education_Level + QC + Percieved_Difficulty + Overshooter + Age + Pre_Conf_A + Pre_Conf_B + Pre_Conf_C

8.4.4.9. Homogeneity of variances

By looking at the table below, we can see that the assumption of homogeneity of variances is met $(p > 0.05)$

Table 8-12: Levene's test

Table 8-12 shows the output for the Levene´s test of equality of error variances. By looking at the p-values, find no significant difference in the variances for either stock. Hence the homogeneity of variance assumption is met.

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
SHIFT A	1,065	152	194	0,338
SHIFT B	1,027	152	194	0,429
SHIFT C	0,963	152	194	0,595

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + AdviceSource + Gender + Education_Level + QC + Percieved_Difficulty + Overshooter + Age + Pre_Conf_A + Pre_Conf_B + Pre_Conf_C

8.4.5. Output two-way mixed MANOVA

Table 8-13: Output two-way mixed MANOVA

Dependent variables are the pre and post-advice confidence level for each stock. The within subjects independent variable is time (before and after advice), while the between subjects independent variable (condition) is the source of the advice, either a financial advisor or a robo-advisor respectively.

Multivariate Tests[a]

Effect			Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Between Subjects	Intercept	Pillai's Trace	0,870	768,772[b]	3,000	345,000	0,000	0,870
		Wilks' Lambda	0,130	768,772[b]	3,000	345,000	0,000	0,870
		Hotelling's Trace	6,685	768,772[b]	3,000	345,000	0,000	0,870
		Roy's Largest Root	6,685	768,772[b]	3,000	345,000	0,000	0,870
	Condition	Pillai's Trace	0,002	,201[b]	3,000	345,000	0,896	0,002
		Wilks' Lambda	0,998	,201[b]	3,000	345,000	0,896	0,002
		Hotelling's Trace	0,002	,201[b]	3,000	345,000	0,896	0,002
		Roy's Largest Root	0,002	,201[b]	3,000	345,000	0,896	0,002
Within Subjects	Advice (time)	Pillai's Trace	0,292	47,343[b]	3,000	345,000	0,000	0,292
		Wilks' Lambda	0,708	47,343[b]	3,000	345,000	0,000	0,292
		Hotelling's Trace	0,412	47,343[b]	3,000	345,000	0,000	0,292
		Roy's Largest Root	0,412	47,343[b]	3,000	345,000	0,000	0,292
	Advice * Condition	Pillai's Trace	0,004	,455[b]	3,000	345,000	0,714	0,004
		Wilks' Lambda	0,996	,455[b]	3,000	345,000	0,714	0,004
		Hotelling's Trace	0,004	,455[b]	3,000	345,000	0,714	0,004
		Roy's Largest Root	0,004	,455[b]	3,000	345,000	0,714	0,004

a. Design: Intercept + Condition
Within Subjects Design: Advice (time)

b. Exact statistic

8.4.6. Assumptions two-way mixed MANOVA

8.4.6.1. Linearity between the dependent variables for each group the independent variable

There was a linear relationship between the pre- and post-confidence levels for all stocks in each condition, as assessed by the scatterplots below.

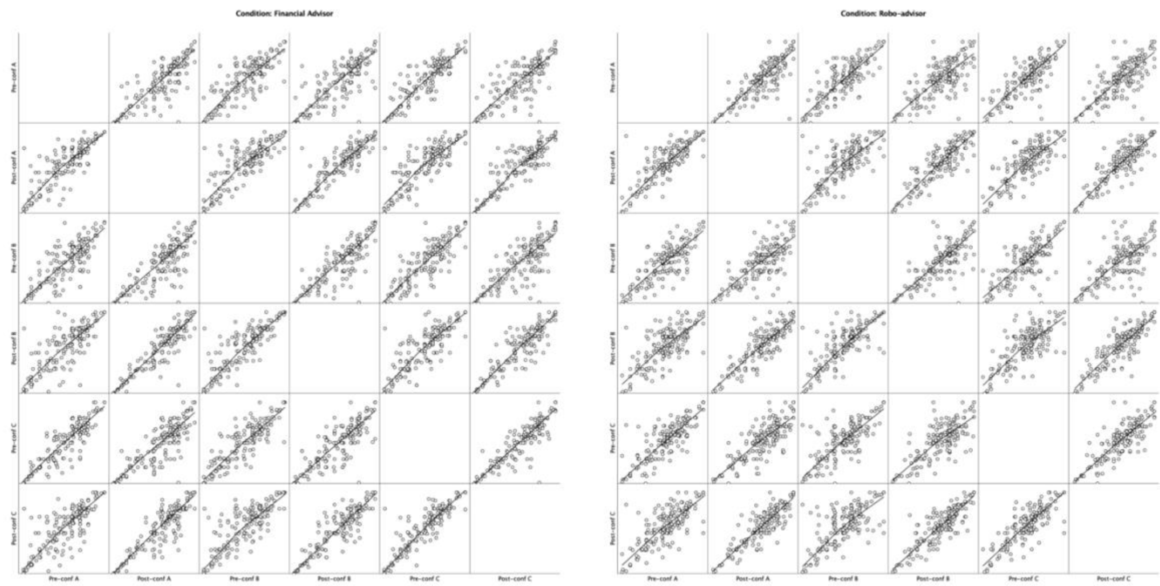


Figure 8-16: Scatterplot Matrix pre and post-conf levels

Figure 8-16 illustrates a scatterplot matrix of the pre and post-advice confidence levels for each stock in each condition. By looking at the lines in each scatterplot, we see that there was alinear relationship between the confidence levels for all stocks in each condition.

8.4.6.2. No multicollinearity - Violated

There was some evidence of multicollinearity in the financial advisor condition. Specifically, the correlation between the post-advice confidence level of 'Stock A' and 'Stock B' were 0.902. The Pearson correlation between all other dependent variables were above 0.90 and did not violate the assumption of multicollinearity (Tabachnick & Fidell, 2012).

Table 8-14: Pearson Correlation Matrix

Table 8-14 shows the Pearson correlation between the SHIFT-scores in each condition. There was evidence of multicollinearity as the Person correlations between 'Stock A' and 'Stock B' were 0.902.

Condition			Pre-advice Confidence Stock A	Post-advice Confidence Stock A	Pre-advice Confidence Stock B	Post-advice Confidence Stock B	Pre-advice Confidence Stock C	Post-advice Confidence Stock c
Financial Advisor	Pre-advice Confidence Stock A	Pearson Correlation	1	,821**	,810**	,776**	,881**	,782**
		Sig. (2-tailed)		0,000	0,000	0,000	0,000	0,000
		N	169	168	169	169	169	169
	Post-advice Confidence Stock A	Pearson Correlation	,821**	1	,797**	,902**	,819**	,879**
		Sig. (2-tailed)	0,000		0,000	0,000	0,000	0,000
		N	168	168	168	168	168	168
	Pre-advice Confidence Stock B	Pearson Correlation	,810**	,797**	1	,845**	,853**	,796**
		Sig. (2-tailed)	0,000	0,000		0,000	0,000	0,000
		N	169	168	169	169	169	169
	Post-advice Confidence Stock B	Pearson Correlation	,776**	,902**	,845**	1	,810**	,891**
		Sig. (2-tailed)	0,000	0,000	0,000		0,000	0,000
		N	169	168	169	169	169	169
	Pre-advice Confidence Stock C	Pearson Correlation	,881**	,819**	,853**	,810**	1	,884**
		Sig. (2-tailed)	0,000	0,000	0,000	0,000		0,000
		N	169	168	169	169	169	169
	Post-advice Confidence Stock c	Pearson Correlation	,782**	,879**	,796**	,891**	,884**	1
		Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	
		N	169	168	169	169	169	169
Robo-advisor	Pre-advice Confidence Stock A	Pearson Correlation	1	,822**	,808**	,736**	,807**	,750**
		Sig. (2-tailed)		0,000	0,000	0,000	0,000	0,000
		N	181	181	181	181	181	181
	Post-advice Confidence Stock A	Pearson Correlation	,822**	1	,723**	,870**	,794**	,871**
		Sig. (2-tailed)	0,000		0,000	0,000	0,000	0,000
		N	181	181	181	181	181	181
	Pre-advice Confidence Stock B	Pearson Correlation	,808**	,723**	1	,766**	,730**	,668**
		Sig. (2-tailed)	0,000	0,000		0,000	0,000	0,000
		N	181	181	181	181	181	181
	Post-advice Confidence Stock B	Pearson Correlation	,736**	,870**	,766**	1	,750**	,847**
		Sig. (2-tailed)	0,000	0,000	0,000		0,000	0,000
		N	181	181	181	181	181	181
	Pre-advice Confidence Stock C	Pearson Correlation	,807**	,794**	,730**	,750**	1	,821**
		Sig. (2-tailed)	0,000	0,000	0,000	0,000		0,000
		N	181	181	181	181	181	181
	Post-advice Confidence Stock c	Pearson Correlation	,750**	,871**	,668**	,847**	,821**	1
		Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	
		N	181	181	181	181	181	181

**. Correlation is significant at the 0.01 level (2-tailed).

8.4.6.3. No univariate outliers – Violated

There were outliers in both conditions as assessed by the boxplots below. Hence, the assumption of univariate outliers was violated.

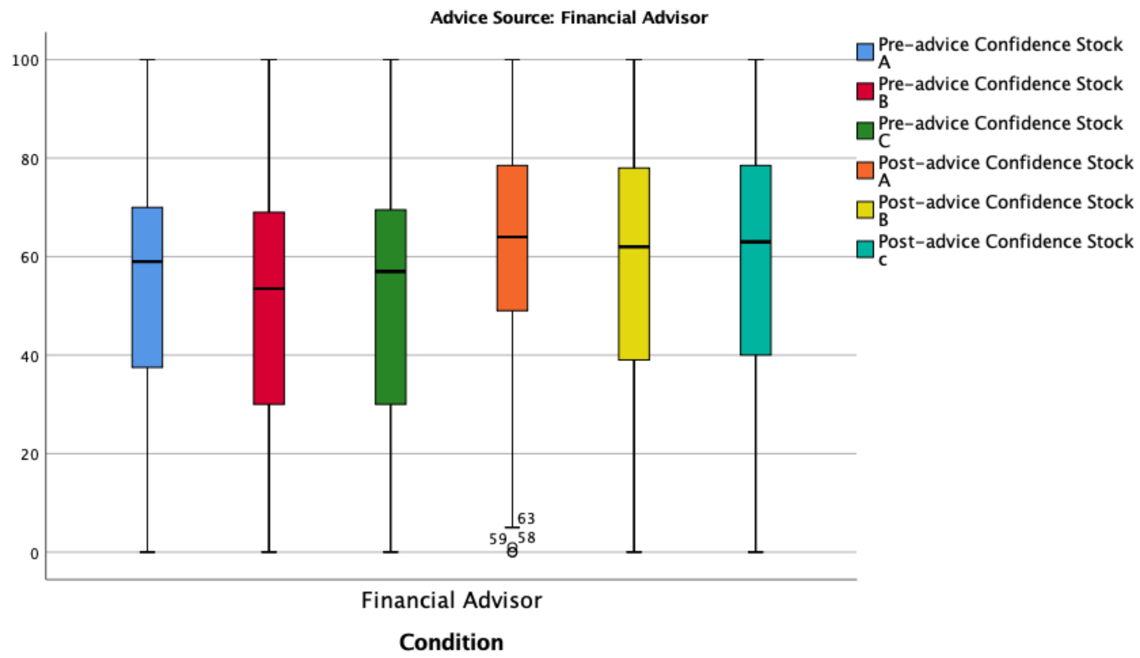


Figure 8-17: Boxplots confidence levels financial advisor condition

Figure 8-17 shows a boxplot of the pre and post-advice confidence levels for the financial advisor condition. By assessing the boxplots, we can see that there were univariate outliers in the data and that the assumption of univariate outliers was violated.

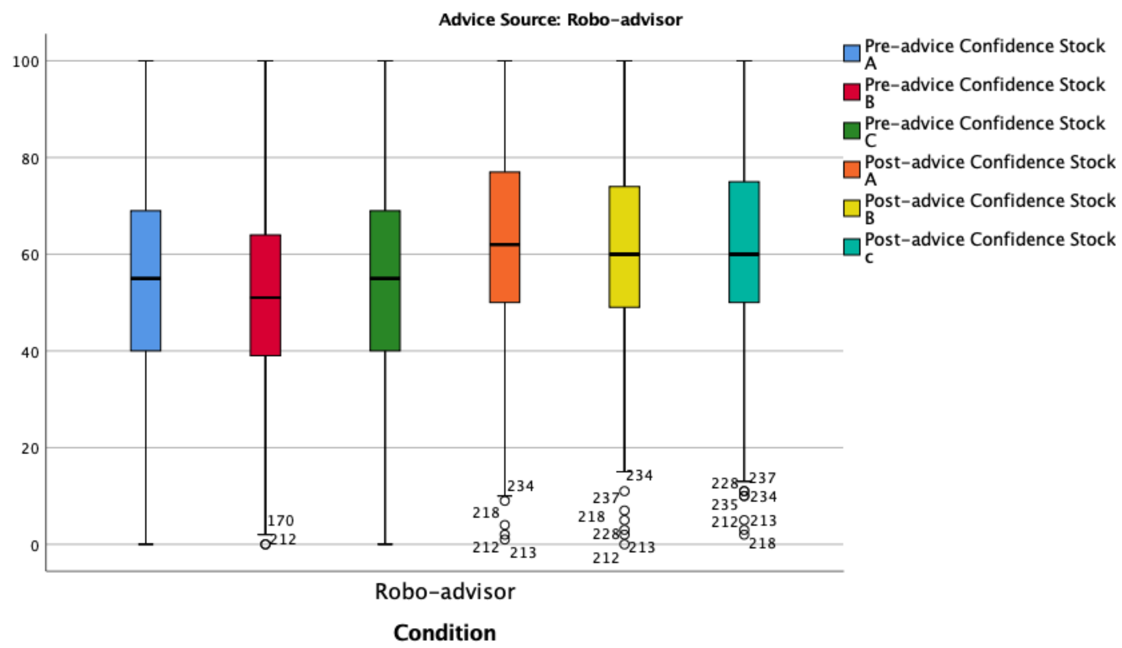


Figure 8-18: Boxplots confidence levels robo-advisor condition

Figure 8-18 shows a boxplot of the pre and post-advice confidence levels for the robo-advisor condition. By assessing the boxplots, we can see that there were univariate outliers in the data and that the assumption of univariate outliers was violated.

8.4.6.4. No multivariate outliers – Violated

The maximum observed Mahalanobis distance was 51.45. The critical value for Mahalanobis distance with six dependent variables is 22.46. Hence, there were five multivariate outliers in the data.

8.4.6.5. Multivariate normality – Violated

The pre- and post-confidence levels was not normally distributed in any condition as assessed by both Kolmogorov-Smirnov's test (all $ps < 0.05$) and a visual inspection of the Q-Q Plots below.

Table 8-15: Test of Normality pre and post-confidence levels

Table 8-15 shows the output of the normality tests for each condition. None of the pre or post-advice confidence levels were normality distributed in any of the conditions (all $ps < 0.05$).

Tests of Normality

Condition		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Pre-advice Confidence Stock A	Financial Advisor	0,124	168	0,000	0,961	168	0,000
	Robo-advisor	0,107	181	0,000	0,982	181	0,021
Pre-advice Confidence Stock B	Financial Advisor	0,086	168	0,004	0,972	168	0,002
	Robo-advisor	0,090	181	0,001	0,985	181	0,049
Pre-advice Confidence Stock C	Financial Advisor	0,111	168	0,000	0,967	168	0,001
	Robo-advisor	0,098	181	0,000	0,981	181	0,013
Post-advice Confidence Stock A	Financial Advisor	0,113	168	0,000	0,941	168	0,000
	Robo-advisor	0,088	181	0,002	0,971	181	0,001
Post-advice Confidence Stock B	Financial Advisor	0,121	168	0,000	0,953	168	0,000
	Robo-advisor	0,095	181	0,000	0,974	181	0,002
Post-advice Confidence Stock c	Financial Advisor	0,115	168	0,000	0,953	168	0,000
	Robo-advisor	0,102	181	0,000	0,980	181	0,010

a. Lilliefors Significance Correction

Q-Q plots financial advisor condition

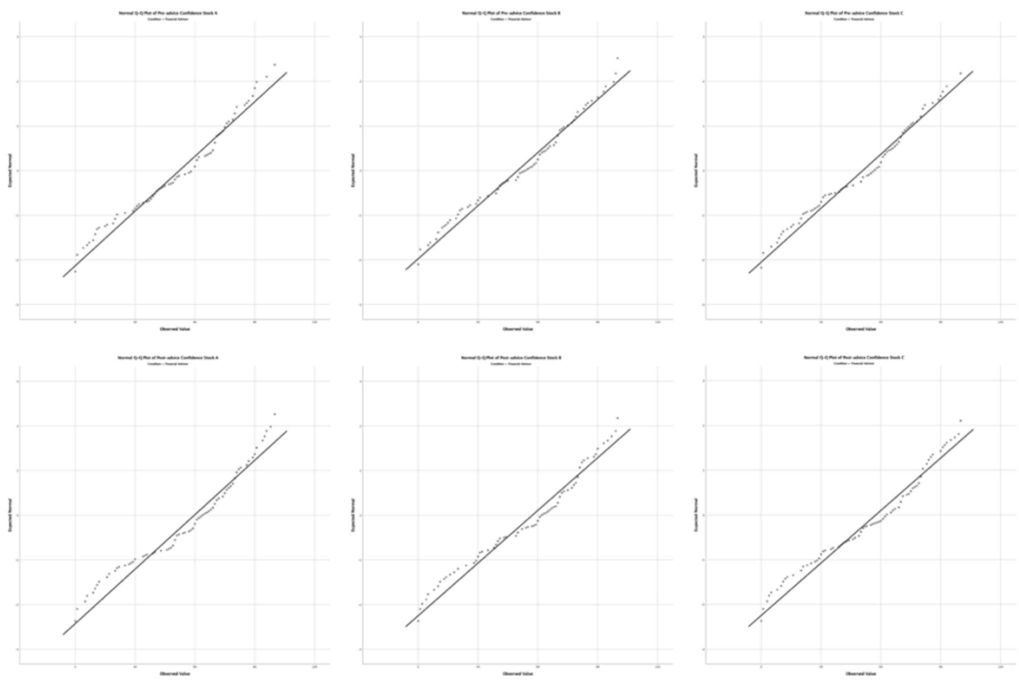


Figure 8-19: Q-Q-plots confidence levels financial advisor condition

Figure 8-19 illustrates a Q-Q plot of the pre and post-advice confidence levels the financial advisor condition. None of the pre and post-advice confidence levels were normally distributed, as illustrated by the deviations from the diagonal line in each plot.

Q-Q plots robo-advisor condition

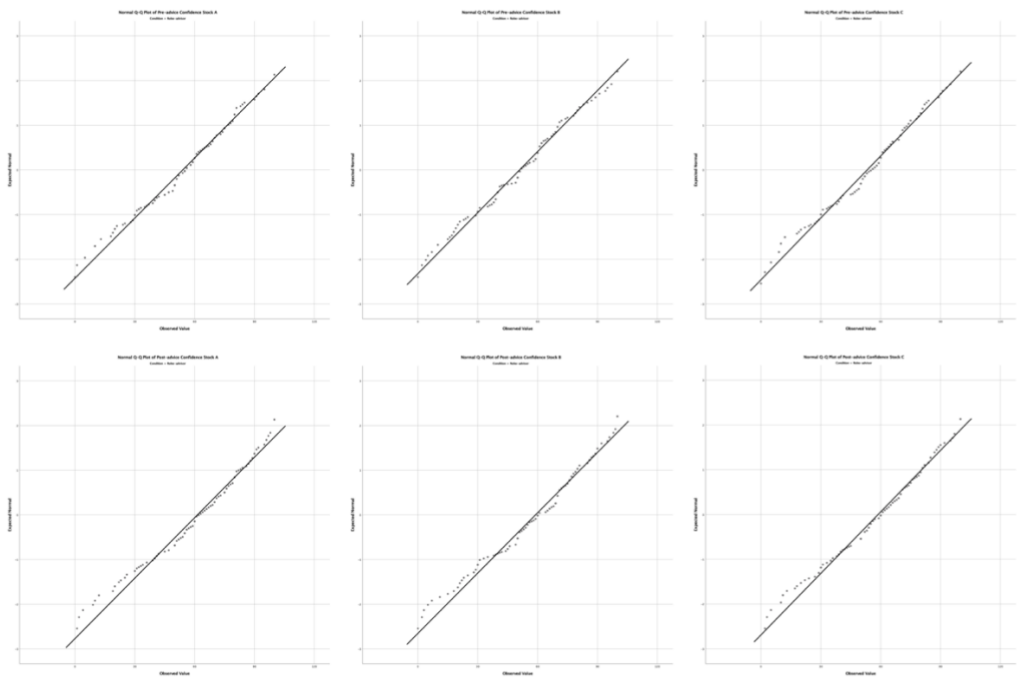


Figure 8-20: Q-Q plots confidence levels robo-advisor condition

Figure 8-20 illustrates a Q-Q plot of the pre and post-advice confidence levels the robo-advisor condition. None of the pre and post-advice confidence levels were normally distributed, as illustrated by the deviations from the diagonal line in each plot.

8.4.6.6. Homogeneity of variance-covariance

By looking at the table below, we can see that the assumption of homogeneity of variance-covariance matrices were met ($p = 0.010$).

Table 8-16: Box's M test

Table 8-16 shows the output for the Box's test of equality of covariance matrices. By looking at the p-value, we find no significant difference in the covariance matrices in the two conditions.

Box's Test of Equality of Covariance Matrices[a]

Box's M	39,590
F	1,851
df1	21
df2	437720,638
Sig.	0,010

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + Condtion
Within Subjects Design: Advice (time)

8.4.6.7. Homogeneity of variances - Violated

By analyzing the Levene's Test of Homogeneity of Variance below, we can see that there are heterogenous variances for some of the confidence levels. Thus, the homogeneity of variances assumption was violated.

Table 8-17: Levene´s Test

Table 8-17 shows the output for the Levene´s test of equality of error variances. By looking at the p-values, we can see that there were heterogeneous variances and therefore violates the assumption of homogeneity of variances

Levene's Test of Equality of Error Variances[a]

		Levene Statistic	df1	df2	Sig.
Pre-advice Confidence Stock A	Based on Mean	4,069	1	347	0,044
	Based on Median	2,794	1	347	0,096
	Based on Median and with adjusted df	2,794	1	342,823	0,096
	Based on trimmed mean	3,959	1	347	0,047
Post-advice Confidence Stock A	Based on Mean	1,836	1	347	0,176
	Based on Median	1,323	1	347	0,251
	Based on Median and with adjusted df	1,323	1	339,421	0,251
	Based on trimmed mean	1,606	1	347	0,206
Pre-advice Confidence Stock B	Based on Mean	6,527	1	347	0,011
	Based on Median	5,734	1	347	0,017
	Based on Median and with adjusted df	5,734	1	345,094	0,017
	Based on trimmed mean	6,413	1	347	0,012
Post-advice Confidence Stock B	Based on Mean	5,369	1	347	0,021
	Based on Median	3,484	1	347	0,063
	Based on Median and with adjusted df	3,484	1	339,181	0,063
	Based on trimmed mean	4,946	1	347	0,027
Pre-advice Confidence Stock C	Based on Mean	6,513	1	347	0,011
	Based on Median	4,661	1	347	0,032
	Based on Median and with adjusted df	4,661	1	341,247	0,032
	Based on trimmed mean	6,415	1	347	0,012
Post-advice Confidence Stock c	Based on Mean	7,426	1	347	0,007
	Based on Median	5,235	1	347	0,023
	Based on Median and with adjusted df	5,235	1	337,945	0,023
	Based on trimmed mean	7,067	1	347	0,008

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Condition
Within Subjects Design: Advice (time)

8.4.7. Output Hotelling's T^2: Winsorized SHIFT-scores

Table 8-18: Output Hotelling's T2: Winsorized SHIFT-scores

Dependent variables are SHIFT Stock A, SHIFT Stock B and SHIFT Stock C. The independent variable (condition) is the source of the advice, either a financial advisor or a robo-advisor respectively. SHIFT-scores above 1.00 were winsorized at 1.00, while scores below 0 were winsorized at 0. Using this method of adjusting SHIFT-scores did not change the result that the participants utilized the advice from a financial advisor and a robo-advisor equally.

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	0,803	464,989[b]	3,000	342,000	0,000	0,803
	Wilks' Lambda	0,197	464,989[b]	3,000	342,000	0,000	0,803
	Hotelling's Trace	4,079	464,989[b]	3,000	342,000	0,000	0,803
	Roy's Largest Root	4,079	464,989[b]	3,000	342,000	0,000	0,803
Condition	Pillai's Trace	0,012	1,345[b]	3,000	342,000	0,260	0,012
	Wilks' Lambda	0,988	1,345[b]	3,000	342,000	0,260	0,012
	Hotelling's Trace	0,012	1,345[b]	3,000	342,000	0,260	0,012
	Roy's Largest Root	0,012	1,345[b]	3,000	342,000	0,260	0,012

a. Design: Intercept + Condition

b. Exact statistic

8.4.8. Output one-way MANCOVA: Winsorized SHIFT-scores

Table 8-19: Output one-way MANCOVA, Winsorized SHIFT-scores

Dependent variables are SHIFT Stock A, SHIFT Stock B and SHIFT Stock C. SHIFT-scores above 1.00 were winsorized at 1.00, while scores below 0 were winsorized at 0. The independent variable (condition) is the source of the advice, either a financial advisor or a robo-advisor respectively.

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	0,142	17,867[b]	3,000	324,000	0,000	0,142
	Wilks' Lambda	0,858	17,867[b]	3,000	324,000	0,000	0,142
	Hotelling's Trace	0,165	17,867[b]	3,000	324,000	0,000	0,142
	Roy's Largest Root	0,165	17,867[b]	3,000	324,000	0,000	0,142
Condition	Pillai's Trace	0,017	1,861[b]	3,000	324,000	0,136	0,017
	Wilks' Lambda	0,983	1,861[b]	3,000	324,000	0,136	0,017
	Hotelling's Trace	0,017	1,861[b]	3,000	324,000	0,136	0,017
	Roy's Largest Root	0,017	1,861[b]	3,000	324,000	0,136	0,017
Gender	Pillai's Trace	0,016	0,858	6,000	650,000	0,526	0,008
	Wilks' Lambda	0,984	,858[b]	6,000	648,000	0,526	0,008
	Hotelling's Trace	0,016	0,857	6,000	646,000	0,526	0,008
	Roy's Largest Root	0,015	1,583[c]	3,000	325,000	0,193	0,014
Education level	Pillai's Trace	0,030	0,825	12,000	978,000	0,625	0,010
	Wilks' Lambda	0,970	0,821	12,000	857,515	0,628	0,010
	Hotelling's Trace	0,030	0,818	12,000	968,000	0,632	0,010
	Roy's Largest Root	0,017	1,376[c]	4,000	326,000	0,242	0,017
Comprehension questions	Pillai's Trace	0,027	2,954[b]	3,000	324,000	0,033	0,027
	Wilks' Lambda	0,973	2,954[b]	3,000	324,000	0,033	0,027
	Hotelling's Trace	0,027	2,954[b]	3,000	324,000	0,033	0,027
	Roy's Largest Root	0,027	2,954[b]	3,000	324,000	0,033	0,027
Perceived difficulty	Pillai's Trace	0,046	0,842	18,000	978,000	0,651	0,015
	Wilks' Lambda	0,955	0,841	18,000	916,896	0,652	0,015
	Hotelling's Trace	0,047	0,840	18,000	968,000	0,654	0,015
	Roy's Largest Root	0,028	1,498[c]	6,000	326,000	0,178	0,027
Overshooter	Pillai's Trace	0,100	12,056[b]	3,000	324,000	0,000	0,100
	Wilks' Lambda	0,900	12,056[b]	3,000	324,000	0,000	0,100
	Hotelling's Trace	0,112	12,056[b]	3,000	324,000	0,000	0,100
	Roy's Largest Root	0,112	12,056[b]	3,000	324,000	0,000	0,100
Age	Pillai's Trace	0,018	1,988[b]	3,000	324,000	0,116	0,018
	Wilks' Lambda	0,982	1,988[b]	3,000	324,000	0,116	0,018
	Hotelling's Trace	0,018	1,988[b]	3,000	324,000	0,116	0,018
	Roy's Largest Root	0,018	1,988[b]	3,000	324,000	0,116	0,018
Pre-Confidence Stock A	Pillai's Trace	0,020	2,242[b]	3,000	324,000	0,083	0,020
	Wilks' Lambda	0,980	2,242[b]	3,000	324,000	0,083	0,020
	Hotelling's Trace	0,021	2,242[b]	3,000	324,000	0,083	0,020
	Roy's Largest Root	0,021	2,242[b]	3,000	324,000	0,083	0,020
Pre-Confidence Stock B	Pillai's Trace	0,008	,828[b]	3,000	324,000	0,479	0,008
	Wilks' Lambda	0,992	,828[b]	3,000	324,000	0,479	0,008
	Hotelling's Trace	0,008	,828[b]	3,000	324,000	0,479	0,008
	Roy's Largest Root	0,008	,828[b]	3,000	324,000	0,479	0,008
Pre-Confidence Stock C	Pillai's Trace	0,023	2,593[b]	3,000	324,000	0,053	0,023
	Wilks' Lambda	0,977	2,593[b]	3,000	324,000	0,053	0,023
	Hotelling's Trace	0,024	2,593[b]	3,000	324,000	0,053	0,023
	Roy's Largest Root	0,024	2,593[b]	3,000	324,000	0,053	0,023

a. Design: Intercept + Condition + Gender + Education_Level + QC + Percieved_Difficulty + Overshooter + Age + Pre_Conf_A + Pre_Conf_B + Pre_Conf_C

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

8.4.9. Output Mann-Whitney U: Winsorized SHIFT-scores

Table 8-20: Output Mann-Whitney U: Winsorized SHIFT-scores

Table 8-20 shows the result of four Mann-Whitney U tests (one for each stock and one for all stocks combined) to test differences in SHIFT-scores between the two conditions (financial advisor versus robo-advisor). SHIFT-scores above 1.00 were winsorized at 1.00, while scores below 0 were winsorized at 0. Using this method of adjusting SHIFT-scores did not change the result that the participants utilized the advice from a financial advisor and a robo-advisor equally.

	Financial Advisor		Robo-advisor		Financial Advisor vs. Robo-advisor		
	Mean Rank	Sum of ranks	Mean Rank	Sum of ranks	U-value	z-value	p-value
Stock A	167.42	27 791.50	171.11	32 239.50	13 930.50	- 1.090	0.270
Stock B	175.87	29 722.00	175.15	31 703.00	15 232.00	- 0.070	0.950
Stock C	182.78	30 890.50	168.70	30 534.50	14 063.50	- 1.310	0.190
Combined	525.39	264 797.50	521.74	282 783.50	135 630.00	- 0.196	0.844

* p < 0.05, ** p < 0.01, *** p < 0.001

8.5. Appendix E: Descriptive statistics 'Study 2'

8.5.1. Descriptive statistics confidence level

Table 8-21: Descriptive statistics confidence levels by condition

Table 8-21 shows the descriptive statistics of the pre and post-advice confidence levels for each stock by condition.

		Financial Advisor		Robo-advisor	
		Mean	SD	Mean	SD
Stock A	Pre advice	52.08	24.51	54.05	22.30
	Post advice	59.72	24.60	61.39	22.16
	% point change	7.48	14.70	7.34	13.27
Stock B	Pre advice	49.73	25.17	50.97	21.95
	Post advice	57.23	25.36	58.94	22.22
	% point change	7.50	14.07	7.97	15.10
Stock C	Pre advice	50.70	24.88	53.25	21.66
	Post advice	57.33	25.55	58.69	21.84
	% point change	6.63	12.14	5.44	13.01
Mean	Pre advice	50.83	24.83	52.76	21.97
	Post advice	58.09	25.15	59.67	22.07
	% point change	7.20	13.66	6.92	13.84