UNIVERSITY OF TRENTO

Doctoral School of Social Sciences

Doctoral programme in Economics and Management

Human technology interaction:

Financial decision making and delegation to algorithms

a dissertation submitted in partial fulfillment of the requirements for the Doctoral degree (Ph.D.) doctoral programme in Economics and Management

Zilia
Ismagilova

2023

Supervisor

- Prof. Matteo Ploner, Università degli Studi di Trento

Doctoral Committee

- Prof. Luigi Mittone, Università degli Studi di Trento
- Prof. Simone Quercia, Università di Verona
- Prof. Caterina Giannetti, Università di Pisa

Abstract

This doctoral thesis consists of three essays within the field of human technology interaction examined through the lens of behavioural and experimental economics. The three essays in this thesis represent three strands helping to reveal the issue of human-machine interaction from different angles.

The first essay contributes to human-machine relations by addressing the problem associated with the problem of an individual experiencing a relative lack of resources that affects human judgment and decision-making in the financial domain. This chapter discusses how policy can leverage emerging technologies to design specific choice architecture that may support more risk-aware decision-making of vulnerable socioeconomic groups. Furthermore, it discusses how behavioural policy initiatives aimed at helping resource-deprived individuals conduct more optimal financial decision-making might be effectively assisted by recent Artificial Intelligence (AI) developments and the associated ethical considerations.

The primary focus of the second essay relates to individual decision-making in a risky environment with algorithm help. By conducting an online experiment, it investigates how humans cognitively offload tasks to algorithms in a risky environment with different time constraints. Results demonstrate that the presence of an AI assistant is beneficial for decision-making only when its accuracy is high.

The third essay continues the investigation of human-technology interactions. The primary attention is paid to how information about the result of the action taken by a human affects the incentive behaviour, depending on the interacting partner.

The main focus concerns how the information about the result (outcome) of the investment affects the reward and punishment behaviour of the participants that interact with Human and Algorithm agents. Specifically, I conduct an experiment investigating the interaction between outcome bias and human/algorithm responsibility.

Keywords: decision-making, SES, AI, nudging, choice architecture, online experiment, algorithm, shift blame, delegation decisions.

Acknowledgements

I have always been obsessed with the idea of studying an individual's behaviour and motives. Critical thinking lived in me from early childhood: I questioned everything and sought to look at everyday things from a different angle. I always looked beyond the standard answers that had to be taken for granted. Perhaps my interest in people brought me to the economics department in Trento. Trento became my home, and the University of Trento my alma mater. Here I began to study the interaction of people and the economy, first at the master's level and then I continued my studies as a PhD student.

PhD opened a new chapter in my life. I am deeply grateful and ny special words are dedicated to my advisor professor Matteo Ploner for his guidance, support, involvement, and help in programming and conducting experiments, feedback on my work, invaluable zoom calls, and, importantly, for his understanding and ability to be a wise mentor throughout my journey. I can not imagine my entire life at PhD without you. Under your supervision, I have grown both as a researcher and as a person.

I would like to express my sincere gratitude to Prof. Matthias Uhl (Technische Hochschule Ingolstadt, Germany), Prof. Caterina Giannetti (Università di Pisa, Italy) and members of my doctoral committee for their valuable comments. Their input has greatly contributed to the improvement of my research, and this dissertation has greatly benefited from their expertise and insights.

I am very grateful to Professor Emanuele Taufer, who supervised me during my master's thesis, for introducing me to the PhD path and bringing me into the world of science. The first year was full of courses, seminars, presentations, and exams. The end of my first year was also tough due to the COVID-19 pandemic. I realized it was a challenging year for everyone on our planet. The pandemic has deprived me of the opportunity to travel and attend summer schools and workshops since they were cancelled due to the epidemiological situation. Nevertheless, I kept fighting and attended online summer school for Behavioural science and even won a special prize that covered my participation in this school. Most importantly, with my colleagues, whom I met at the summer school, I won the opportunity to publish my article in a special journal issue. Continuing to work half remotely and half in presence, I attended a few conferences, among which was an Italian Chapter of the Association for Information Systems, where I presented my second paper. Thanks to professor Paolo Barbieri for sponsoring an online experiment that made that publication possible. To my pleasure, the conference was held in Trento, and I am very grateful to all the organizers and the administration staff.

I thank the University of Trento for funding my scholarship and the opportunity to be part of the academic community, for organizing seminars, conferences and workshops; the School of Social Sciences, professors and faculty members (especially prof. Luigi Mittone, Enrico Zaninotto, Fabio Zona, Lucia Savadori, Luciano Andreozzi, Ericka Costa, Carlo Fezzi, Maria Laura Frigotto), and administrative staff (especially Nicole Bertotti, Davide Santuari, and Stefania Valentini) for providing the facilities, assistance and support, organizing seminars, and facilitating professional development. Many thanks to all the colleagues, conference attendees, CEEL lab members, and the doctoral committee for their valuable comments, which helped to improve my research.

I want to thank my husband Ivan and fur friend Harvey, who have emotionally supported me through my journey. The PhD thesis project is hard work, requiring great dedication, self-discipline, effort and perseverance. Figuratively speaking, you are the captain and the only crew member on a ship ploughing the oceans of the unknown. This experience changed me,

brought me to an entirely new level of understanding myself and the world, and gave me opportunities I had not imagined. Now, at the end of it, I am sure I would have done this path again. Thanks to my resources, I plan to continue to grow and develop to contribute to improving the beautiful world of mankind.

Introduction

The rapid development of technology has an indelible impact on human life. Technologies cover the diverse aspects of human life, reshaping human behaviour and creating new forms of interaction between individuals and he technological realm. Technologies impact society through their products and processes, affecting the quality of life and how people act and interact. Digital transformation also changed how technology contributes to the decision-making process and brings to humans new possibilities, like delegating specific tasks to facilitate decision making. A human increasingly relies on the results of the performance of a machine, results of computational capabilities, forecasting and predicting possible outcomes.

In this thesis, the goal is to explore diverse perspectives on how to aid and improve the interaction between humans and machines; how to support a human being in making more optimal decision-making, including financial choices, depending on the specific conditions of both the human himself and the environment in which he is. The general idea of the thesis revolves around humans and their interaction with technology, influenced by various factors such as time constraints, penalties, and a tendency to offload or delegate tasks to algorithms and smart assistants.

In answering whether human-machine interaction differs depending on social factors, the first chapter of the thesis aims to explore human-machine interaction, looking at the problems of the individual experiencing relative resource scarcity (which is often linked with low socioeconomic status) affecting human judgment and decision-making abilities in financial domains. Relative resource scarcity, in the form of economic inequality, is rising worldwide and is a problem with enormous societal costs. Governments spend a considerable amount of money on reducing childhood poverty and investing in public poverty services, including, for example, loss of economic productivity and increased health and crime costs. The first chapter suggests avenues where policy can leverage emerging technologies to design specific choice architecture that may support more risk-aware decision-making of vulnerable socioeconomic groups. It is essential to improve the decision-making process for socially and economically vulnerable individuals through utilising tools and techniques presented by technological advances. Additionally, this thesis explores how behavioural policy initiatives aimed at helping resource-deprived individuals to make more optimal financial decision-making might be effectively assisted by recent Artificial Intelligence (AI) developments and the associated ethical considerations. This chapter contributes to an increased understanding of the psychological mechanisms involved in decision-making under resource scarcity and how anomalies in such decision-making strategies might be better mitigated by the use of AI in order to help resource-deprived individuals achieve better life outcomes.

Building on the topic of human-machine interaction, this thesis endeavours to partially address the following question - what factors contribute to and diminish the possible benefit of algorithm aid? The second essay further focuses on individual decision-making within a risky context, with algorithmic support. Given the ongoing fourth industrial revolution, algorithmic assistance has now become an indispensable aspect of modern individuals' daily lives. The increasing practice of outsourcing not only physical activities but also mental and cognitive aspects has amplified the influence and significance of algorithmic support. Nevertheless, since technology mainly relies on a fixed set of rules or algorithms, being AI-augmented, people remain the main player (who are not always aligned with mathematical rules and rationality), who monitor performance and make a final decision. Hence, it is vital to investigate the human-machine interaction, find the factors influencing human decision-makers, and help to identify the most optimal solutions. Consequently, a better understanding of the interaction between human decision-making and AI support might help create choice architectures that would aid human problem solvers in hybrid intelligent systems, provide solutions to businesses through optimising processes, and help the government and policymakers design strategies.

As stated earlier, thoroughly analysing the interaction between humans and machines is paramount. Does the inferential and judgemental capacity of an individual within a specific domain vary based on their interaction with a particular partner? It is crucial to explore not only how the interaction of a person and the algorithm affects the decision-making process but also how this interaction will affect the consequences and judgements of humans, including the validity of the decision made. Meanwhile, people tend to evaluate decisions after the fact, meaning that evaluation of the quality of the decision made includes information not only about decision strategy but also the result of the outcome. Moreover, in real-life settings, various factors need to be carefully controlled, and hence, the quality of the outcome is not necessarily a reflection of the quality of decision-making. Consequently, humans tend to rely more on the outcome or result rather than on the strategy of the decision. Thus, this thesis explores the phenomenon of outcome bias, which refers to the tendency of individuals to rely more on information about the outcome when judging decision quality rather than the decision-making process. The third essay investigates the human technology interaction from the point of view of how information about the result of the action taken affects the incentive behaviour of a human, depending on the interacting partner. Furthermore, the main focus concerns how the information about the result (outcome) of the investment affects the reward and punishment behaviour of the participants that interact with Human and Algorithm agents.

The present thesis makes substantial contributions to the ever-growing and dynamic field of research on human-machine interactions, providing a comprehensive analysis of this subject matter from three distinct perspectives. Firstly, this work focuses on the impact of algorithms on specific social strata, with a particular emphasis on how these technologies can aid and support vulnerable populations. Secondly, this thesis investigates the effects of critical factors such as time constraints and penalties on the decision-making processes that occur within algorithmic systems. Finally, this study examines the complex processes of decision delegation between humans and algorithms, with a specific focus on the domain of financial decision-making. Taken together, these three lines of inquiry make valuable contributions to our understanding of the multifaceted complexities that characterise human-machine relationships.

Contents

1 How can AI Technologies Aid Financial Decision-Making of People with Low Socioeconomic Status? 1

1.1 Introduction 1

1.2 Literature review 3

1.2.1 Socioeconomic status 3

1.2.2 Impact of SES on decision-making 4

1.2.3 Digitisation and the impact on decision-making . . . 5

1.2.4 The digital divide 5

1.2.5 Behaviour change techniques (BCTs) and nudging as possible solutions 7

1.2.6 The role of AI in decision-making 8

1.3 The effects of low SES on cognitive development and decision-making 8

1.3.1 The effect of resource scarcity on decision-making under uncertainty 12

1.4 How can technology aid financial decision-making for low SES individuals? 14

1.5 Potential risks of using AI in financial decision-making and ethical concerns 19

1.6 Discussion and conclusion 22

2 Does the penalty and time pressure affect offloading? 25

2.1 Introduction . . . 25
2.2 Literature review . . . 28
2.3 Method and Materials . . . 32
2.3.1 The Task . . . 32
2.3.2 Treatments . . . 33
2.3.3 Design . . . 34
2.3.4 Procedures and Participants . . . 35
2.4 Research Hypotheses . . . 36
2.5 Analysis . . . 38
2.5.1 Impact of AI on performance . . . 38
2.5.2 Impact of time constraints on performance . . . 40
2.5.3 Impact of Penalty on Performance . . . 42
2.5.4 Impact of AI Accuracy on performance . . . 44
2.5.5 Survey . . . 49
2.6 Discussion . . . 50
2.6.1 Limitations . . . 52
2.7 Conclusions . . . 52

3 Ain't blaming you: Delegation of financial decisions to humans and algorithms 55

3.1 Introduction . . . 55
3.1.1 General framework and motivation. . . . 59
3.1.2 Do choices made by a machine shift attention from the outcome to the process? . . . 61
3.1.3 Shifting responsibility and delegation of decisions. . . . 63
3.1.4 Risk-taking, if a human makes a decision on her own, would she act riskier or safer? . . . 66
3.2 Research Hypotheses . . . 67
3.3 Method and Materials . . . 69

3.3.1 Task 69
3.4 Treatments 71
3.4.1 Preliminary session 71
3.4.2 Information (between-subjects) 72
3.4.3 Reward Before (Ex-ante) [Exp 1] 72
3.4.4 Reward After (Ex-post) [Exp 2] 72
3.4.5 Procedures and Participants 73
3.4.6 Players 74
3.4.7 Payoff 74
3.4.8 Survey 75
3.5 Research Hypotheses 75
3.6 Analysis 77
3.6.1 Result 1. 77
3.6.2 Result 2. 81
3.6.3 Result 3. 84
3.6.4 Result 4. 85
3.7 Discussion 87
3.8 Conclusion 90

Conclusions **92**

Appendices **95**

A **95**

A.1 Survey 95
A.2 Survey responses before and after the experiment 96
A.3 Second part. 97
A.3.1 Instructions for the second part 97
A.3.2 Training 98
A.4 Third part 99

A.4.1 Survey 2 99

B **101**

B.1 Questionnaire 101

Bibliography **106**

List of Tables

2.1 Table of treatments. 34
2.2 Descriptive statistics for nAI and AI sessions: performance and timing. 39
2.3 Performance depending on Time constraints and AI/nAI presence. 40
2.4 Performance by penalty and assistance. 43
2.5 Performance regression. 45
2.6 Reaction time regression. 47

3.1 Investment levels 70
3.2 Dimensions that are subject to manipulation in the experiment. 72
3.3 Experiment participants structure 74
3.4 Ex-ante and Ex-post investment 77
3.5 Ex-ante and Ex-post investment - Wilcoxon test 78
3.6 Reward by Ex-post|Success, Ex-ante, Ex-post|fail regression 79
3.7 Principal reward in Ex-post Success and Ex-post Fail regression 80
3.8 Principal reward in Ex-post|Success and Ex-post|Fail regression 82
3.9 Principal Reward in Ex-post|Success and Ex-post|Fail regression by investment level 83

A.1 The probability of the correct suggestions 97

List of Figures

2.1 Example of two-dimensional stimuli. 33
2.2 Example of AI assistance. 34
2.3 Frequency of correct answers distribution, nAI vs. AI 39
2.4 Performance by time constraint and assistance. 41
2.5 Performance by time constraint and assistance in rounds. . . 42
2.6 Timing by round and block for each treatment and session. . 43
2.7 Performance by penalty and assistance in rounds. 44

3.1 General experiment design 71
3.2 Rewards in Ex-post|Success, Ex-ante, and Ex-post|Fail . . . 78
3.3 Rewards in Ex-post|Success, Ex-ante, and Ex-post|Fail depending on investment level 80
3.4 Reward of Algo and human in Ex-post and Ex-ante 81
3.5 Principal reward in Ex-post Success and Ex-post Fail by investment level 83
3.6 Delegation choices in Ex-ante and Ex-post 84
3.7 Investment level in Ex-post and Ex-ante 86

A.1 Survey responses before and after the experiment 96
A.2 Example of a training task 98
A.3 Example of a feedback on training task 99

B.1 No delegation to Algo 103

B.2 Agent delegate to Algo 103
B.3 Principals answer. No delegation to Algo 104
B.4 Principals answer. Agent delegate to Algo 104
B.5 Delegation by Agent, age less or equal to 40 105
B.6 Delegation by Agent, age more 40 105

Chapter 1

How can AI Technologies Aid Financial Decision-Making of People with Low Socioeconomic Status?

1.1 Introduction

In this chapter, the aim is to provide an overview of the vulnerability of individuals of low socioeconomic status to choices that undermine their well-being in the face of economic decisions that require a sufficient understanding of risk and suggest avenues where policy can leverage emerging technologies to design specific choice architecture that may support more risk-aware decision-making of vulnerable socioeconomic groups. The primary motivation underlying this chapter is the desire to improve the decision-making process for socially and economically vulnerable individuals through utilising tools and techniques presented by cutting-edge tech-

Reported in this chapter is the result of a joint work with Christian T Elbæk, Ifeatu Uzodinma, and Panagiotis Mitkidis.

nological advancements.

Relative resource scarcity, in the form of economic inequality, is rising worldwide. Recent reports from the OECD have outlined the severity of the problem by showing that an increasing number of people living in developed economies such as the US and Europe are slipping into lower-income classes (OECD, 2019). Notably, rising inequality and poverty are problems with enormous societal costs. In 2015, childhood poverty cost the US 5.4 per cent of its GDP, amounting to $1.03 trillion (McLaughlin & Rank, 2018), and from 2016 to 2017, the UK spent £78 billion just on public poverty service costs (McCarthy, 2016) including, for example, loss of economic productivity and increased health and crime costs (McLaughlin & Rank, 2018). Therefore, issues of poverty and income inequality keep defining political agendas worldwide, as predicted in 2013 by former US president Barack Obama terming it "the defining issue of our time" (Sargent, 2013).

Motivated by this, the current study conducts a thorough and critical examination of the current literature on how individual experiences of relative resource scarcity (specifically, low socioeconomic status) may have an effect on human judgement and decision-making in financial domains. Based on these findings, I discuss how behavioural policy initiatives aimed at helping resource-deprived individuals conduct recent technological developments and the associated ethical considerations might effectively assist more optimal financial decision-making. The main focus of the discussion would be related to how novel technologies, specifically Artificial Intelligence (AI), can aid in improving financial decision-making for individuals with low-risk awareness. The present study proposes possible avenues where policymakers can leverage emerging AI technologies to design specific choice architecture that may support more risk-aware decision-making of vulnerable socioeconomic groups.

Further, this chapter will also address concerns related to criticisms of implementing nudging on population and the ethics of utilising nudges in vulnerable populations particularly. There are two prominent critics - the most crucial critic autonomy violation and the second one related to concerns that the population should be manipulated by government activity. Even though Sunstein pointed "It is pointless to object to choice architecture or nudging as such" (Sunstein, 2015) because it's simply not possible to avoid it, there are still many disputes.

The current chapter contributes to an increased understanding of the cognitive processes related to decision-making in the context of limited resources. Additionally, this study attempts to mitigate anomalies in these decision-making strategies by utilising advanced artificial intelligence (AI) techniques, with the ultimate goal of enhancing the life outcomes of individuals who are deprived of resources.

1.2 Literature review

1.2.1 Socioeconomic status

Generally, individuals from lower socioeconomic classes (SES), defined as individuals with low household income, educational level and occupational security, are over-represented in several worrying statistics (Baker, 2014). Lack of income equality and access to education and medicine are considered as main factors to a disparity in SES. People with low SES have higher obesity rates (Drewnowski & Specter, 2004), lower levels of education (West, 2007), higher rates of teenage pregnancy (Young, Martin, Young, & Ting, 2001), take on more debt (Hartfree & Collard, 2014), consume more alcohol (Khan, Murray, & Barnes, 2002), and gamble more than people in higher income brackets (Blalock, Just, & Simon, 2007). These findings have one thing in common: successful decision-making in these specific domains requires the individual to be able to focus attention, resist stimuli, and delay gratification. This is a central problem for people with low SES as empirical evidence has identified that individuals who do not have enough of a needed resource discount the future and fail to focus on the outcome that would serve them best (Mullainathan & Shafir, 2014; Shah, Mullainathan, & Shafir, 2012).

1.2.2 Impact of SES on decision-making

Subjective Expected Utility is a theoretical framework that may help recognise the possible mechanisms at play and target the problem mentioned above. The theory has dominated economic theory on choice in decision environments characterised by imperfect information (Camerer & Weber, 1992) and delineates how economic agents respond to uncertainty about states of nature by subjectively assigning probabilities to alternate outcomes in the absence of complete information (Savage, 1972). Intuitively, this subjectivity suggests heterogeneous beliefs about the future across economic agents, with recent studies showing that people in the low SES demography consistently discount future pay-offs more than high SES individuals (Oshri et al., 2019; Carvalho, Meier, & Wang, 2016).

It might be that low SES individuals generally hold pessimistic beliefs about unknown future states of nature. For them, having experienced events associated with low SES, such as frequent adverse income shocks (Haushofer & Shapiro, 2013), current consumption is preferred to some unknown future (Amir, Jordan, & Rand, 2018). With the tendency for this demographic group to be comparatively more risk-averse and less willing to invest in education (Haushofer & Fehr, 2014), they are likely to be more vulnerable to financial illiteracy, and financial exclusion (Barboni, Cassar, Demont, et al., 2017). In an economic decision environment, the absence of requisite financial literacy could indicate low SES individuals' inadequacies in risk cognition. Without satisfactory awareness and comprehension of risk, low SES individuals will likely defer to their inherently high level of risk aversion and inordinately discount pay-offs in the future. This failure to identify better long-term outcomes can, in turn, lead to a series of consistently poor financial decisions that make it near impossible for these individuals to escape poverty (Carvalho et al., 2016; Haushofer & Fehr, 2014).

1.2.3 Digitisation and the impact on decision-making

As financial decision-making becomes increasingly digitised with a growing number of interactions happening online (Accenture, 2019), financial institutions such as banks, pension funds, and mortgage lenders have rapidly adopted new digital technologies to offer services entirely online (Gomber, Kauffman, Parker, & Weber, 2018). On one side, this means that data collection becomes highly personalised. Currently, algorithms predict the probability of certain product purchases and customers' willingness to pay for them. Since more data allow businesses to view better consumers' willingness to pay, big data for commercial companies is a tool to find the perfect target for their product at a given time. Generally, this might lead to a situation when big data facilitates price discrimination because of companies' advantage of the information and the opportunity to set discriminatory prices. However, scholars suggest that personalised pricing is also beneficial for consumers (Dubé & Misra, 2017).

1.2.4 The digital divide

On the other side, embraced with big data, humans might become overwhelmed with data. This can lead individuals to become victims of information isolation by their initial digital choices, perpetuated through search history, location, and past click behaviour known as *filter bubbles* (Pariser, 2011). Although the phenomenon is more often associated with search engines, the formation of civil opinion, and marketing promotions, it is not clear yet how these filter bubbles affect the behaviour of economic agents, especially those making financial decisions with a high level of uncertainty. Overloaded information-rich environments, governed chiefly by artificial engines, interfere with humans' ability to embrace the existing information and consequently make optimal decisions. While such a data-rich digitised environment gives the feeling of control, conscious and optimal choice, individuals might fall into self-deception and lose awareness in such environments, especially if they lack the ability to discount future outcomes and focus on the task at hand (Helbing, 2019; Lipina & Posner, 2012; Shah et al., 2012; Banerjee & Mullainathan, 2008). This might, in turn, lead to polarisation among social groups, ultimately deteriorating the already decreased economic and societal position of individuals with low SES. At the same time, a considerable portion of the population does not have access to digital technologies and is therefore limited from information and resources. Consequently, they have the disadvantage of access to products of information and communication technologies such as digital e-commerce, online education, employment, and other digital benefits, which is known as *digital divide* (Ragnedda & Muschert, 2013). The lack of access to technology and the notion of *digital divide* is driven by a lack of finance, education and cultural resources (Livingstone & Helsper, 2007). Consequently, this can lead to a wider gap between the low and higher SES population. This situation is like a loop from which it is difficult to escape.

1.2.5 Behaviour change techniques (BCTs) and nudging as possible solutions

As a possible solution, techniques to change behaviour for vulnerable populations can help avoid the digital gap and help bridge the gap between low and higher SES individuals. One of the well-known instruments to change individuals' behaviour is behaviour change techniques (BCTs) (Lyons, Lewis, Mayrsohn, & Rowland, 2014). Furthermore, the number of included techniques is less important than the combination and quality of implementation because an application with fewer but more effective techniques has more impact on humans. Goal-setting, rewards, feedback, and self-monitoring of behaviour are essential to behavioural intervention. Including goal setting and providing information about consequences in the app is essential. How effective they are compared to human financial assistants is a big question. However, not everyone can afford a human financial assistant and his services. However, there is a need to use Assistants carefully to avoid harming the one who uses them. Otherwise, it might cause provoke inadequate activity. While BTCs mainly address health issues, in the financial domain, if we plan to use rising digital technologies through applications to support low SES individuals, the application to perform successfully should include behaviour change techniques. There are positive examples of correcting assistance. For example, (Fanning, Mullen, McAuley, et al., 2012) provides some preliminary support for mobile technology interventions to increase physical activity behaviour.

As suggested in this thesis, the concept of gentle correction without choice restriction and compulsion is better known as "nudging" (Thaler & Sunstein, 2009). While there are a lot of debates and critics of the utilisation of the nudging (Bovens, 2009; Wilkinson, 2013), it is still considered a valuable tool to build a prosocial choice architecture and to reduce disparities among the population. Furthermore, (Mrkva, Posner, Reeck, & Johnson, 2021) demonstrated that “good nudges” designed to increase the selection of superior options reduced choice disparities, improving choices more among individuals with lower SES, lower financial literacy, and lower numeracy than among those with higher levels of these variables.

1.2.6 The role of AI in decision-making

To avoid deplorable consequences, especially for low SES individuals, this chapter suggests using novel technologies, specifically AI, to improve individual choices in complicated decision-making environments. While the literature suggests various definitions of AI, in this chapter, I define AI as a system’s ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation (Kaplan & Haenlein, 2019). Notably, the idea of improving individual choices through the implementation of such technological tools should be organised without restricting options. Suggested implementations aimed to aid in improving financial decision-making for individuals with low-risk comprehension because these AI technologies (like machine learning), thanks to their ability to tackle issues involving large data sets, can take into account the current financial limitations of an agent’s personal financial situation and hence make the optimal financial choices more salient to such agents, without restricting their number of choices.

1.3 The effects of low SES on cognitive development and decision-making

A large and rapidly expanding body of research in the neuro and cognitive sciences has produced evidence demonstrating that growing up and living with a low SES can have detrimental effects on the development of particular vital cognitive functions of the human brain (Duval et al., 2017; Hackman, Farah, & Meaney, 2010; Giedd et al., 1999). These cognitive functions include areas of the brain associated with inhibitory and interference control, cognitive flexibility, stimuli control, and focus regulation, generally known under the broad term *executive functions* (Diamond, 2013; Sarsour et al., 2011; Hackman et al., 2010). Low SES is also directly tied to structural differences in the brain of children, in areas of the brain that are linked to educational skills and achievements (Hair, Hanson, Wolfe, & Pollak, 2015), and has been shown to be associated with adult earnings and the number of working hours in later life (Duncan, Ziol-Guest, & Kalil, 2010). IQ variance in low SES families is even shown to be prominently explained by the shared environment. At the same time, such a relationship does not exist for affluent families, where IQ variance is extensively explained by genetics (Turkheimer, Haley, Waldron, d'Onofrio, & Gottesman, 2003).

As a result of this suppressed development in cognitive functions, living with resource scarcity, characteristic of low SES, has been shown to reduce what is often conceptualized as "mental bandwidth" (Mani, Mullainathan, Shafir, & Zhao, 2013). This means that scarcity makes individuals experience shifts in their cognitive attention or focus regulation (Tomm & Zhao, 2016) that can lead to sub-optimal economic decisions because, in the economical choice environment, specific problems which are considered less critical or distal are neglected while others considered more proximal receive more attention (Shah et al., 2012; Spears, 2011). For example, an attention shift may result in undesirable behaviour in the form of impulsive decision-making, where short-term gains, i.e. consumption, are valued higher than the long-term ones, such as investing in education that should typically increase the economic agent's welfare (Zhao & Tomm, 2018). Simply inducing thoughts about finances has been shown to impede the cognitive function of poor participants. A similar effect was not found among the "rich" subjects in the study, indicative of a resource scarcity effect on cognitive ability (Mani et al., 2013). More empirical findings support this as individuals from low-income US households have been shown to be more present biased in intertemporal choices when decisions are made just before payday (Carvalho et al., 2016). Scholars argue that such findings might explain why specific economic problems, such as over-borrowing, are more prevalent in resource-scarce populations (Shah et al., 2012). Individuals from low SES backgrounds tend to act more impulsively, exhibit greater risk-taking behaviour, and approach temptations faster (Griskevicius et al., 2013).

Overall, these findings indicate that resource scarcity makes people focus on the problems at hand while neglecting the more long-term outcomes of their behaviour (Hall, Zhao, & Shafir, 2014; Mullainathan & Shafir, 2014; Shah et al., 2012). Experiencing scarcity is associated with reduced behavioural control, leading to poorer short-term economic decision making, with detrimental consequences for the long-term well-being of resource-deprived individuals (Spears, 2011). Because the deprived individuals' focus on regaining resources in the short-term overshadows the opportunity to achieve better prospective outcomes, such opportunities are simply favoured less compared to the immediate relief of deprivation (Shah, Zhao, Mullainathan, & Shafir, 2018; Shah et al., 2012; Spears, 2011). Shah et al. (2018) experimentally induced emotions relating to distinct correlates of low SES environments to isolate empirical effects of poverty on cognitive functioning. The first study examined how often induced concerns about money made subjects think about cost-related items –compared across high and low SES participants. Findings revealed that low-income individuals were more likely to think about cost-related items. In the second experiment, participants were primed with a treatable life-threatening health experience and asked to write down the three most salient words that came to mind. Written words were grouped into 'emotion-related' or 'money-related.' In the results, low SES individuals wrote down more money-related words than high SES individuals. The third experiment specifically investigated interference regulation. The instructions required participants to allow their minds to wander freely while actively suppressing any thoughts related to monetary costs. Findings revealed that low SES subjects had more intrusions by cost-related thoughts than high SES subjects –indicative of the former being less able to regulate intrusions and maintain focus than the latter.

Haushofer and Fehr (2019) attempted to distinguish the effects of negative income shocks. In a lab experiment, subjects were randomly assigned different starting endowments to experimentally create 'rich' and 'poor' subject groups. Both groups were then given tasks to complete to earn cash, after which all participants were exposed to positive and negative income shocks. Their findings revealed increased discounting resulting from the negative income shocks, though this effect was determined to be consistent across participants in both groups with non-significance in the discrepancy between the 'rich' and 'poor' groups. Importantly, this evidenced lack of behavioural control can lead to the agents becoming confined to so-called "poverty traps" – where the consequences of decision-making aimed at restoring resources in the short-term generate a vicious circle of having to engage in additional risky economic decision-making to alleviate the consequences of previous economic choices; for instance, in the form of borrowing money at high-interest rates to pay off current debts (Gandy, King, Streeter Hurle, Bustin, & Glazebrook, 2016).

The importance of this problem in regards to policy is highlighted by longitudinal research showing that individuals who grow up in families with low SES are much more likely to end up with low SES in adulthood as well

(Lesner, 2018; Wagmiller & Adelman, 2009), indicating that poverty is transmitted intergenerationally (De Lannoy, Leibbrandt, & Frame, 2015). Thus, the detrimental consequences of economic decision-making under scarcity are tied not only to the individual's long-term well-being but also to the individual's family and, thus, future generations. This underscores the importance of implementing targeted policy campaigns aimed at helping individuals with low SES exhibit more optimal economic decision-making, not only for the life outcomes of the current generation but also for the well-being of future generations (Gandy et al., 2016).

1.3.1 The effect of resource scarcity on decision-making under uncertainty

The neoclassical economic theory assumed the goal of human decision-making to be utility maximization (Jerčić et al., 2012; Camerer & Weber, 1992). Behavioural theory has since found ample evidence to contradict this, establishing subjectivity as the core of human decision-making, which presents as heterogeneous agents subjectively assigning probabilities to the likelihood of occurrence across alternative outcomes (Loewenstein, Nagin, & Paternoster, 1997; Camerer & Weber, 1992; Kahneman & Tversky, 2013). These subjective probability assignments and expectations result from an agent's subjective perceptions and are often derived from experience (Fisher, Hull, & Holtz, 1956). Specific experiences, especially among low SES individuals, have been linked to imprudent spending (Sheehy-Skeffington, 2020; Amir et al., 2018), and negative experience (e.g., change in economic circumstances) has been shown to induce negative affective states like stress (Haushofer & Shapiro, 2013), anxiety and unhappiness (Ozer, Fernald, Weber, Flynn, & VanderWeele, 2011); all having an adverse effect on time-discounting and revealed preferences (Haushofer &

Fehr, 2014). The prior adverse economic experience thus makes an agent prone to mistakes in financial decision-making (Carvalho et al., 2016), so where the choice environment is identical to both low and high SES individuals, the decision responses of low SES individuals are biased by emotions from the recall of previous adverse events. Emotions, broadly defined by their valence (negative or positive) and their intensity (level of arousal) therefore play a significant role in human decision-making (Jerčić et al., 2012; Loewenstein, 2000). In a cold state (little or no arousal), an agent's emotion-informed response is more controlled or reflective as opposed to the reverse hot state (a heightened state of arousal), where the individual exhibits automatic responses and less control over their behaviour.

However, not all instances of arousal and decision-making driven by emotion necessarily lead to negative behaviour and sub-optimal payoff. Seo and Barrett (2007) show that making decisions based on emotions can lead to a positive payoff in some instances. In theory, emotions are exposed to bilateral effects that may lead to biased choices that are detrimental to the agent's well-being or reflective responses that lead to optimal decision-making (Jerčić et al., 2012). It follows that the ability to regulate emotional responses can lead to improved decision-making, especially in stochastic environments with imperfect information (Heilman, Crişan, Houser, Miclea, & Miu, 2010). This is critical for low SES individuals who, by their demographic features, generally have less education and, therefore, less chance of understanding stochastic environments. Research has shown that emotions have a strong impact even among highly trained and risk-savvy traders; trading loss is usually followed by high-risk aversion, and extreme caution (Fenton-O'Creevy, Soane, Nicholson, & Willman, 2011). It is conceivable, therefore, that both low SES and high SES individuals make choices in similar decision environments but with fundamentally dissimilar appraisals of the presented choice architecture. The level of risk does not vary between the two groups (Haushofer & Fehr, 2014). However, exogenous conditions like frequent income shocks, limited access to credit, and low financial literacy vary the level of risk perception. As a result, the poor will consistently exhibit higher present bias than the non-poor (Muraven & Baumeister, 2000).

1.4 How can technology aid financial decision-making for low SES individuals?

The current chapter outlines a novel interdisciplinary approach to understanding and combating the fundamental problem of how to better help resource-deprived individuals through specialised behavioural policy initiatives, an issue of prime importance for researchers in economic policy and policy-makers alike. In this chapter, I suggest that institutions adopt specific policy initiatives to develop selected AI technologies as "nudging tools" to help individuals experiencing relative resource scarcity make more optimal economic decisions that can improve individual welfare and reduce societal costs associated with poverty.

Following the previous section, can adverse economic decision outcomes of low SES demographic groups be mitigated? The situation is exacerbated by the information-rich environment where agents are increasingly surrounded by information that fits their initial interests while ignoring other relevant data. Choice manipulations and filter bubbles are not necessarily harmful to rational agents, but as established by science, *homo economicus* is rarely observed in everyday life (Sunstein, 2018; Thaler, 2000). Commercial companies, therefore, often use filter bubbles and information traps to manipulate individual decision-making. By using AI, these companies use collected information, for example, to create personalised marketing campaigns and advertisements based on personal preferences, behaviour, and beliefs (Hern, 2014). Promotions are centred on using an individual's past behaviour in connection with subconscious decision-making biases (social influence, emotional motivations, scarcity, etc.) to manipulate consumer choices (Dowling et al., 2020; Parker & Lehmann, 2011; Cialdini & Cialdini, 2007; Taylor, 2000). This consequently means that individuals might not necessarily be aware of the reasons triggering their actions.

While this is a general problem, low SES individuals are more vulnerable to these manipulations due to their reduced attention span, which leads such individuals to underestimate risk, discount the future, and favour short-term outcomes to restore needed resources (Schmidt, Neyse, & Aleknonyte, 2019; Tomm & Zhao, 2016; Hall et al., 2014; Mullainathan & Shafir, 2014; Griskevicius et al., 2013; Griskevicius, Tybur, Delton, & Robertson, 2011; Sarsour et al., 2011; Hackman & Farah, 2009), something that could have severe consequences for the individuals and the society more generally. Furthermore, Botta and Wiedemann (2020) note that companies can bring potential benefits to certain customer segments and contribute to the redistribution of wealth between different categories of consumers. Bergemann et al. (2015) showed that additional (personal) information could increase and decrease consumer surplus. In this respect, "strategic" customers, concerned by the risks of price discrimination, can estimate the value of their personal data and, therefore, hide their identity during online activity (Acquisti & Varian, 2005). Instead, "myopic" consumers (i.e., digitally illiterate) would be less cautious about exposing their private data on the Internet and, therefore, would be more vulnerable to potential price discriminations (Acquisti & Varian, 2005). However, Bellaflamme and Vergote (2016) showed that less cautious customers might benefit from price discrimination even in a monopolistic scenario: the customers relying on anonymising technologies would be subject to the uniform price, which might be higher than a personalised one (Belleflamme & Vergote, 2016).

While scientists warn about the potentially harmful effects of the development of full AI, where robotic intelligence supersedes that of humans (Cellan-Jones, 2014), this essay suggest a change of emphasis to look at the existing (weak) AI with a modern perspective, in line with new trends in technology adoption and in particular with the new concepts of augmented intelligence and its function in society. While augmented intelligence is an umbrella term that takes on a certain sense depending on the context (Porter, 2017; Pasquinelli, 2015), in this essay, it is defined as the use of technology to expand human information processing capabilities (Sharma, 2019). As high technology is penetrating society, policy-makers can benefit from the opportunities of digital technologies by combining technological solutions with social norms and legal regulations.

Recently, information systems (IS) have become a significant component in enhancing competitive advantage on an organizational level, supporting decision making, and facilitating day-to-day operations (Checkland & Holwell, 1998). AI is expanding the scope of IS applications not only through task automation but also through integrating and mimicking human intelligence. AI can augment human capability by providing data-driven insights on risky financial decisions at speed, making more optimal choices and reminding individuals of alternative ways to improve their welfare in the long term (Karlan, McConnell, Mullainathan, & Zinman, 2016). Examples of such an application are currently arising in the fintech industry, where an increasing number of AI startups implement solutions aimed at helping individuals with their financial decision-making (Kaya, Schildbach, AG, & Schneider, 2019; Kashyap, 2018; Lui & Lamb, 2018).

Specifically, this essay suggests using these novel advancements in technologies and AI specifically to improve financial decision-making for low SES individuals with low-risk comprehension. While many governments have already implemented behaviourally informed policies using choice architecture (Mousavi, Kheirandish, et al., 2017) to make individuals more environmentally friendly (Slapø & Karevold, 2019; Nielsen et al., 2017; Sunstein, 2016) or to promote retirement savings (Thaler & Benartzi, 2004), this essay mainly focuses on how technology and AI particularly can be used to nudge individuals in digital financial decision environments. In addition to existing policies aimed at helping the poor, this essay proposes to include behaviorally informed technological policies for the personal banking sector. Resource scarcity can also stimulate the development of nudge interventions. An increasing number of people have to use banking services from online agents and hence interact through automated online support systems (Accenture, 2017). Managing finances in an environment swamped with information, such as lengthy contracts and difficult-to-understand banking terms, can be challenging for any decision-maker but especially for individuals who lack focus regulation and capacity for assessment of financial risk, which are some of the cognitive characteristics associated with low SES (Mullainathan & Shafir, 2014; Mani et al., 2013). Hence, this makes it difficult to make informed choices for low SES individuals. In the present context, nudging can serve as a powerful tool for promoting desired outcomes. One approach to nudging is to simplify well-being information, which can be accomplished through the strategic use of plain language, visually intuitive materials, and simplified instructions. By leveraging these techniques, it might be possible to facilitate comprehension and promote more efficient outcomes. Nudges can promote financially literate behaviour, for example, by incorporating push reminders and recommendations within mobile applications. Such nudges can help individuals become more conscious of their financial choices and encourage them to adopt habits that are more aligned with financial literacy principles. More advanced virtual financial assistants integrate with voice assistants (web and mobile) to provide individuals with more convenient banking services, ranging from basic knowledge and support requests to personal finance management and conventional banking. For instance, chat-bots use AI to generate personalized financial real-time advice with budgeting, savings goals, and expense tracking. Based on accumulated data, AI can read and analyze contracts, notify of specific terms, cancel money-wasting subscriptions, and find better insurance options - those activities in which people with low SES can be especially vulnerable due to low financial literacy. Including chat-bots and digital assistants might increase transparency and clarity by analyzing and interpreting massive datasets that are difficult to comprehend, particularly for less educated and financially illiterate individuals (Gnewuch, Morana, & Maedche, 2017). Increasing anthropomorphism might result in an even higher level of users' compliance with a chat-bot's request for service feedback (Adam, Wessel, Benlian, et al., 2020), which, with a well-formulated government policy, leads to more optimal decision making. Therefore, integrated government initiatives, including AI, to interact and communicate with users to make public services more tailored to all groups of individuals must be a prime focus.

Often, low SES individuals are faced with the problem of limited access to loans because banks cannot assess the risk of default. To avoid such discrimination, scholars (Óskarsdóttir, Bravo, Sarraute, Vanthienen, & Baesens, 2019) suggest using AI in assessing the credit-scoring of low SES individuals, using data collected from mobile phones, such as detailed call records, social media analysis, or information on customers' credit and debit accounts. This AI initiative is primarily aimed at the individual's external environment and assists in facilitating access to credit and insurance for low SES individuals. Furthermore, this essay suggests that the internal component is likely associated with decision-making because these tools allow legitimate individuals to develop confidence in their creditworthiness and reduce the variance of risk perception, thereby increasing the confidence and positive attitude they often lack.

Utilizing AI to help individuals with low SES make better financial decisions comes with great individual as well as societal benefits. While the future might seem bright, some of the significant challenges that AI systems currently face are the lack of trust (Davenport, 2019), algorithm avoidance (Castelo, Bos, & Lehmann, 2019), the risk of biases (Frank, Chrysochou, Mitkidis, & Ariely, 2019; Awad et al., 2018) and, more importantly, major regulatory concerns (Buiten, 2019). Establishing new social norms and legal regulations for social intelligence would require sufficient transparency and accountability.

1.5 Potential risks of using AI in financial decision-making and ethical concerns

Autonomy and automation should therefore come with responsibility, hence requiring a legal framework for such technologies. Two conditions should hold to achieve the most beneficial outcome in interactions with technology. Firstly, data used for targeting and enforcing social protection programs should be exhaustive and include all ranges of social and economic layers of the population. Otherwise, in case of the absence of data for certain societal groups, this can lead to discrimination and a larger gap between demographic groups. Additionally, depending on the system's characteristics and particular circumstances, individuals should have a choice – to rely on a system or not. Individuals must be given the possibility to consciously decide for or against a decision or action; otherwise, individual autonomy and responsibility are undermined.

Individual autonomy and responsibility can be undermined in cases if she becomes too reliant on smart technologies without consciously considering the implications of her actions. This might be an often case in interaction with the "hidden" assistants, which are so firmly ensconced in humans' daily life that humans do not even realise how often they physically and mentally rely on technology. The phenomenon of using physical activity to alter the information processing requirements of a task in order to reduce cognitive demand is called *cognitive offloading* (Risko & Gilbert, 2016). Cognitive offloading can be beneficial in some cases though it can result in disaster in others (Parasuraman & Riley, 1997). Cognitive offloading can harm performance or might not be advisable, for instance, in tasks concerning efficiency (Weis & Wiese, 2019b). Therefore, the subject of the impact of AI on human cognitive offloading and its impact on behaviour should be studied in depth before specific policy initiatives are implemented.

The use of AI in the financial domain poses other potential risks that are closely associated with the limitations of AI. Scholars define such potential risks of using AI as embedded bias, black box problem, cyber-security, data privacy, robustness, and impact on financial stability (Boukherouaa, AlAjmi, Deodoro, Farias, & Ravikumar, 2021). Embedded bias refers to biases related to the design and implementation of the algorithms, for example, at the development stage or the training stage of the system. These potential early-stage defects can lead to further system operation distortion. For example, in 2015, Google withdrew their recognition app because it tagged people with darker skin as gorillas, as well as Amazon's case when the system turned out to weed out female candidates because while training algorithm had a biased sample. Hence an incomplete or unrepresentative sample to train the system continues to be a critical issue for designing AI systems. Indeed, scholars recognise that AI systems raise questions and new unique risks to the financial system's integrity and safety, which remains to be assessed (Boukherouaa et al., 2021).

Certainly, the use of nudge tools may raise ethical concerns related to issues of social justice, equality, and autonomy. While nudge tools can be useful in promoting positive behaviour, their effectiveness might vary depending on individual preferences, motivations, and circumstances. Apart from the fact that AI needs to be generally designed intelligently with respect to input and training data, it is also necessary to understand that the distribution of the diversity of users who will use this system is extensive. For example, considering psychology and behavioural responses, a solution suitable for one population category may be doubtful for another category. Another concern related to the use of nudges is the fact that nudges can be seen as a form of coercion or manipulation, as they might be used to push people into certain behaviours or choices without a proper understanding of the consequences, which might be risky for people with low SES people given that they are already more vulnerable to financial illiteracy (Barboni et al., 2017). Designing AI with a bias towards one particular demographic group can lead to prejudice and inequality in society and unintentional discrimination against other groups. Additionally, algorithm-based nudges can disrupt individual autonomy and decision-making processes, which can be seen as a violation of personal freedom. Therefore, when using AI-based nudges for people with low SES, it is critical to consider the ethical implications and ensure that they are applied in a way that respects individual autonomy and is consistent with principles of social justice and equity.

The fundamental idea of choice architecture, nudging, is to improve individual choices in complicated decision-making environments without restricting any options (Hansen et al., 2019; Leonard, 2008). The dark side of a *nudge* is a *sludge*, which directs attention to choices that make the decision-maker worse off, e.g., by encouraging self-defeating behaviour such as taking loans with unfavourable terms when better options exist

(Thaler, 2018). As outlined above, AI is a powerful technology and can be used to simplify and improve financial decision-making under uncertainty for people with low SES. However, it can just as well be used as a *sludge* to guide attention towards choices that will make the decision-maker worse off, ultimately trapping low SES individuals in poverty. Therefore, to avoid (intentionally or unintentionally) undermining individual freedom of decision-making and ethical guiding principles, demanding certain quality standards and sufficient transparency is necessary when utilizing AI to aid the financial decision-making of those less well-off.

Nevertheless, despite the potential risks, scholars argue that AI can serve as a useful tool in minimising human bias in decision making by eliminating irrational biases stemming from the subjective interpretation of data (Mayson, 2018; Silberg & Manyika, 2019) or by mitigating human bias (Miller, 2018).

1.6 Discussion and conclusion

Growing up and living with low SES can have detrimental effects on successful decision-making in financial choice environments characterized by a high level of uncertainty. As outlined in the present chapter, novel technologies, specifically AI, can be utilized to simplify, organize, and optimize these financial environments for individuals who experience a lack of behavioural control and therefore discount the future and fail to focus on the outcome that would serve them best. However, this form of technological *nudging* comes with considerable responsibilities and ethical considerations. Therefore, this essay urges regulators and policy-makers to implement legal guidelines for using AI in financial decision-making so that the outcome can be beneficial for those most in need. While proposing specific legal frameworks and ethical guidelines concerning the use of AI in nudging better financial choices is beyond the scope of this thesis, I acknowledge that this is one of the essential considerations concerning how such technologies should be successfully implemented.

Furthermore, as outlined above, human cognitive offloading and its impact on behaviour should be studied in depth before specific policy initiatives are implemented. Future research across the behavioural sciences should thus aim to comprehensively investigate how specific problems related to financial decision-making under scarcity might be alleviated by using AI and particularly how such support might be done without putting the individual at increased risk. I urge future research to investigate and develop precise, practical implementations of how AI could aid financial decision-making under economic scarcity. This form of research will then benefit not only the ones with the least available resources but society as a whole.

In the following chapters, I focus on human-algorithm interactions and how digitalisation and algorithms and machines influence human behaviour, decision making and beliefs. Digitalisation is an extensive and global phenomenon affecting many domains of society. It is an inevitable phenomenon and will be carried out constantly and accelerating every year as one of the components of the overall progress of civilisation. Since digitalisation has penetrated all spheres of life, and smart systems and artificial intelligence are becoming ubiquitous in everyday actions and decision-making, there is still a lack of research on the impact of smart assistance on the human decision-making process. Given the increasing speed of transactions, the amount of processed daily information, and the number of participants on the ground, enormous pressure is created for the average person, which cannot but affect a person's behaviour and mental processes.

Based on this, questions arise - for example, how factors such as speed and time pressure will affect the interaction between a human and an algorithm. Will the algorithm contribute to faster decision making or slow it down, and how will this affect the "independence of decision making" by a human, whether he will offload or mindlessly follow the "advice" of the algorithm? As well as the question - does the accuracy of an algorithm affect a human's trust in it and also on how often a human will delegate his decisions to the algorithm.

As discussed earlier in this chapter, cognitive offloading represents a very attractive function of humans, especially in conjunction with modern, sophisticated algorithms. A better understanding of the interaction between human decision-making and AI support might help create choice architectures that would aid human problem solvers in hybrid intelligent systems, provide solutions to businesses through optimising processes, and help the government and policymakers design strategies. Hence in the second chapter, I plan to expand the topic of human-algorithm interaction regarding how it affects the offloading process. Particularly, I plan to focus on the effect of the accuracy of the algorithm and the punishment for a wrong decision made on offloading.

i

Chapter 2

Does the penalty and time pressure affect offloading?

2.1 Introduction

Smart assistants are becoming ubiquitous in everyday decision-making. The growth of AI and its impact on human beings have attracted much controversy in recent years. For example, Stephen Hawking claimed that the development of full AI could mark the end of the human race (Cellan-Jones, 2014). On the other hand, the CEO of IBM, Ginni Rometty, argues that AI technologies are designed to augment human intelligence and that the partnership between humans and machines will make humans better (Carpenter, 2015). Nevertheless, the prospect of a proliferation of AI has created social alarm. Consequently, dozens of academic scientists, AI experts, and ethicists have signed an open letter calling for comprehensive research into the social and economic implications of AI technologies. This action is a logical and natural inquiry, especially considering that AI applications continue to spread, and human problem solvers are faced with the impact of AI on daily routines, decision-making processes, and strate-

Reported in this chapter is the result of joint work with Matteo Ploner.

gic choices. In this sense, artificial intelligence becomes an indispensable assistant in daily and business production tasks for a human, facilitating his routine and taking parts of his tasks.

During previous industrial revolutions, human beings have successfully used outsourcing machines to replace people's physical abilities. In light of the ongoing fourth industrial revolution, more than ever, human beings outsource not only physical activity but also mental or cognitive processes. The externalisation of cognitive processes into technological aids is called *Cognitive Offloading* (S. J. Gilbert, 2015; Kirsh & Maglio, 1994; Risko & Gilbert, 2016; Weis & Wiese, 2019a). The availability of smart technologies, such as Artificial Intelligence (AI), and access to the answers "at fingertips" impact how individuals think, offload, and process information (Ward, 2013; Sparrow, Liu, & Wegner, 2011). There is another relatively new concept that has emerged in recent years called "algorithm take-up" which also aims to reduce cognitive load on human. Specifically, algorithm take-up refers to the practice of delegating decision-making processes entirely or partially to algorithmic systems. Thus, algorithm take-up can be viewed as a specific instance of the broader concept of cognitive offloading. However, the current paper will use the term cognitive offloading as a broader concept involving a wider range of strategies for reducing cognitive load via external aids.

Despite ongoing research on the intersection of AI technology and human decision-making, further inquiry is needed to fully understand the impact of smart assistants. The most apparent difficulty of AI-related research lies in the uneven spread of AI technologies at the industrial and individual levels, in the ambiguity of perceptions, lack of trust, and awareness among humans. Consequently, more research needs to be conducted on the impact of AI on human cognitive offloading and its application in behaviour science. It is unclear how humans and AI can complement, augment, or replace each other in decision-making processes and how it will ultimately affect their performance. How does the type of partner — human or AI — affect the human decision-making process, perception of choice, and choice itself? Furthermore, if new technological tools can affect the decision-making process, can scholars and policymakers help people to make better decisions and achieve better results? These questions are crucial because human decision-makers remain responsible for decision outcomes.

Organizational scholars have attempted to explore the complementary of humans and AI in the context of decision making. For example, Davenport and Kirby (Davenport & Kirby, 2016) argue that machines will not be displacing intellectual workers soon. Indeed, there are psychological barriers to this, such as the fear of anthropomorphism, distrust of machines in solving ethical questions, and the functional limitations of algorithms, such as a constraint of performance in accordance with the input data. In addition, Jarrahi (Jarrahi, 2018) suggests that AI systems should be designed to augment, not replace, human contributions.

Current research in behavioural science needs to provide clear and comprehensive empirical evidence of how the decision-making processes are affected in a mixed environment in which humans and AI complement each other and work together in risky, uncertain, or fluid situations. Meanwhile, the application and impact of AI on the economy and society are growing fast. For example, automated decision-making has been gaining ground recently in several domains, such as insurance underwriting and financial trading (Davenport & Kirby, 2016). In the financial sector, approximately 70 per cent of all market transactions are now made by automated trading algorithms (Helbing, 2019). Nevertheless, since technology mainly relies on a fixed set of rules or algorithms, being AI-augmented, people remain the main player (who are not always aligned with mathematical rules and rationality), who monitor performance and make a final decision. Hence, it is vital to investigate the human-machine interaction, find the factors influencing human decision-makers, and help to identify the most optimal solutions. Consequently, a better understanding of the interaction between human decision-making and AI support might help create choice architectures that would aid human problem solvers in hybrid intelligent systems, provide solutions to businesses through optimizing processes, and help the government and policymakers design strategies.

2.2 Literature review

Certainly, decision making is a complex process where individuals used to rely on aids - tools facilitating the process and helping to make a choice. How a successful choice was made ultimately affects the performance. One of the major tasks of technology as an external tool is to aid and augment memory and cognition, in order to optimize decision-making efficacy. Indeed, people tend to offload their cognition to facilitate performance, in many situations, using internal and external sources (Maeda, 2012; Wilson, 2002; Clark & Chalmers, 1998; Kirsh & Maglio, 1994). Without technological support, many humans struggle to solve cognitive tasks involving arithmetic (Osiurak, Navarro, Reynaud, & Thomas, 2018; Walsh & Anderson, 2009), spatial navigation (Fenech, Drews, & Bakdash, 2010), or prospective memory (Cherkaoui & Gilbert, 2017; N. Gilbert & Stoneman, 2015) efficiently. Facing difficult questions make people start thinking about computers, and people expecting to access information in the future, are less likely to remember the information itself and instead better remember where to access it (Sparrow et al., 2011). Scholars note that technology and the internet have become the main form of external memory, where information is stored collectively outside humans.

Although AI-augmented and technology-aided may seem advantageous for decision-making, research says that people are not always willing to use technological aids. For instance firms (Sanders & Manrodt, 2003) and professionals (Fildes & Goodwin, 2007) (Vrieze & Grove, 2009) regularly choose not to use algorithms as their primary forecasting method, which often results in less accurate forecasts (Sanders & Manrodt, 2003). The uncertain nature of the domain, such as driving a car, making medical decisions, predicting socio-political events, and investing, is another example of unwillingness to rely on technology (Dietvorst & Bharti, 2020). This is despite the fact that technological tools might outperform humans in many domains (Dietvorst & Bharti, 2020; Kuncel, Klieger, Connelly, & Ones, 2013; Frey & Osborne, 2017). Such as, humans generally do not trust algorithms in solving ethical issues considering algorithms to be insufficiently experienced in these subjects, so expertise is another impediment to using decision aids because more knowledgeable experts tend to reject recommendations from both algorithmic and human advisors (Logg, Minson, & Moore, 2019; Yaniv, 2004; Arkes, Dawes, & Christensen, 1986).

The other cause affecting technology usage, particularly AI, is trust. The issue of trust in technology is especially acute in the financial environment. Humans tend to trust decision-making to other humans rather than machines, possibly because of the belief that humans are more rational and empathetic compared to machines (Filiz, Judek, Lorenz, & Spiwoks, 2022). Recent literature on Internet banking shows that the lack of trust is considered one of the main reasons why consumers are still reluctant to conduct their financial transactions online (Flavián, Guinaliu, & Torres, 2006; Luarn & Lin, 2005; Mukherjee & Nath, 2003; Rotchanakitumnuai & Speece, 2003). Some studies found that there is a significant increase in users' trust and the user's delegation of controls to autonomous systems as the risk decreases and vice-versa (Perkins, Miller, Hashemi, & Burns, 2010).

Furthermore, there was a significant difference between a user's initial trust before and after interacting with an autonomous system under varying risk conditions (Ajenaghughrure, da Costa Sousa, & Lamas, 2020) and during cooperative interaction tasks aiming to achieve the same goal (Satterfield, Baldwin, de Visser, & Shaw, 2017). Research defined three sources of variability in trust in automation: dispositional, learned, and situational (Hoff & Bashir, 2015). Dispositional factors include the age, culture, and personality of the trustor among other characteristics. Learned trust is based on past experiences relevant to a specific automated system. Finally, situational factors concern the context of the human-automation interaction and various aspects of the task, such as workload. As shown, risk and trust are essential components of human-technology interaction. In this chapter, the work will focus on situational trust and manipulate specific contextual variants.

As mentioned above, scholars attempt to understand decision making and cognition offloading in different contexts and affecting factors in technology-infused environments. Scope of research has touched upon a preferable strategy for cognitive processing (Morgan, Patrick, Waldron, King, & Patrick, 2009; Gray, Sims, Fu, & Schoelles, 2006; Gray & Fu, 2004), offloading of memory and cognition depending on preexisting beliefs (Weis & Wiese, 2019b), high access cost (Walsh & Anderson, 2009; Gray et al., 2006), and difficulty of the cognitive task (Risko & Gilbert, 2016; Risko, Medimorec, Chisholm, & Kingstone, 2014; Walsh & Anderson, 2009). The cunning and proficiency of humans to offload cognition depending on specific goals were demonstrated by Weis and Wiese (Weis & Wiese, 2019a). The authors investigated the influence of different performance goals — maximizing speed or accuracy. By measuring how frequently and how proficiently humans offloaded, the authors showed that participants offloaded less in the speed than in the accuracy goal condition. Although the authors suggest that participants can perform proficiently without external guidance if they have a clear performance goal and stable feedback, it is unclear whether this ability will remain in a risky environment. (Wu, Schulz, Pleskac, & Speekenbrink, 2022) suggests that people make faster, more random decisions when the risks are low but slow down and think longer when the risks are high, thus demonstrating that people are sensitive to the cost-benefit trade-off of increased deliberation. Additionally, it is unclear how a risky environment affects humans' cognitive offloading based on their current goals.

Thus, it is not clear what are the factors and conditions under which individuals would certainly follow or rely on technological aids. Moreover, individuals might exhibit an acute selectivity when they rely upon a recommendation (Liang, Sloane, Donkin, & Newell, 2022), and sometimes they over-rely on technology (Parasuraman & Riley, 1997). Might the performance change under certain conditions depending on the goals set? Will participants show selectivity if their choice in risky conditions affects their own benefit?

Significant factors commonly found in real-life situations that can have a significant impact on human behaviour and influence decision making are (lack of) time and penalty (or punishment), which this article focus on. Time pressure refers to situations where people are required to make decisions quickly, often without access to needed information or resources. Punishment refers to the negative consequences that may result from a particular decision, like loss of investment. One of the main reasons this article focuses on time constraints and penalties is that they are factors that can have a significant impact on decision making. For example, a lack of time can occur when a stockbroker needs to make a quick and beneficial decision in a dynamic market, while punishment can be a concern in the event of losing her own investment and losing a percentage for trading client's funds.

Research about the influence of time constraints and penalties on the decision making and the offloading process can facilitate the development of more effective strategies to mitigate their negative effects and provide more positive effects. Hence, the research question is based on the confrontation of cognitive offloading in risk conditions involving time constraints and penalties.

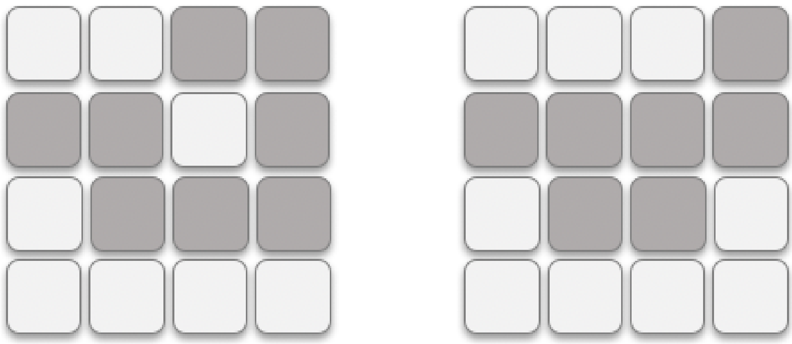
2.3 Method and Materials

The research question addressed in the current paper relates to individual decision-making in a risky environment with the assistance of an algorithm. Specifically, the study investigates how *penalties* and *time constraints* affect offloading behaviours, which result in a different performance when participants interact with AI technologies. Accordingly, the following experiment has been developed to explore this phenomenon.

2.3.1 The Task

As a reference task for the experiment, I take the well-known mental rotation paradigm (Shepard & Metzler, 1971). This task represents a real effort task (Charness, Gneezy, & Henderson, 2018) which captures essential features of human interaction with continuously expanding technologies. Specifically, the task aimed to investigate human decision making in interaction with technology under pressure and the cognitive offloading in the presence of technology - how the interaction might enhance or reduce the cognitive load. Participants are asked to compare pairs of two-dimensional objects on the screen (Fig. 2.1) and solve the task by applying mental resources or AI suggestions and give an answer - whether the objects are identical or different. Within each pair, the stimuli can be rotated around an axis 360 degrees; thus, rotation is not considered a difference. Changing shape, size, and specularity is considered a difference.

Figure 2.1: Example of two-dimensional stimuli.



2.3.2 Treatments

Two factors are experimentally manipulated at once: *Time constraint* and *Penalty*.

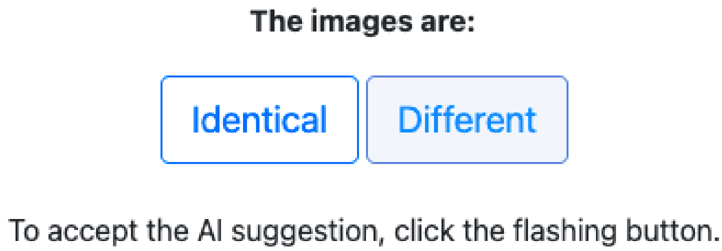
- Time Constraint (between-subjects design).
  - High time constraints (highTime-C) - 10 seconds to solve the task is given;
  - Low time constraints (lowTime-C) - 17 seconds to solve the task is given;
- Penalty (within-subject design) - Risk component.
  - High penalty (highPnlt) - each mistake in a task costs 10 points;
  - Low penalty (lowPnlt) - each mistake in a task costs 1 point;

There was conducted two sessions to have a control group: with AI assistance (AI); and with no AI assistance (nAI). AI assistance is represented by a suggestion if objects are identical or different (Fig. 2.2). Within each pair, the stimuli differ in spatial orientation in a specific amount of degrees in either direction.

Table 2.1: Table of treatments.

Sessions	Time Constraint (between-subjects design)	Penalty (within-subject design)
nAI	High time constraints	High penalty
AI	Low time constraints	Low penalty

Figure 2.2: Example of AI assistance.



2.3.3 Design

The experimental session can be divided into three independent parts.

First part. After accepting the consent form, participants are given a definition of AI, followed by a short survey regarding participants' attitudes towards AI and information technology. Prior to the second part, participants should also take an attentiveness test.

The second part consists of the core experiment task - the visual comparison task. Participants are asked to compare pairs of two-dimensional objects on the screen (Fig. 2.1) and solve the task by applying mental resources or AI suggestions and give an answer - whether the objects are identical or different (see Section 2.3.1). The participants were subjected to either high or low time constraints, whereby the inability to provide a response within the designated time constraint would be treated as an incorrect response and would trigger the participant to progress automatically to the next task. Before the main block of the experiment with tasks, participants were given three trials to play and familiarize themselves with the experiment. Overall, participants face 40 tasks, divided into four rounds of 10 trials each. In each round, the accuracy of the AI support changes and participants are informed of this. Respectively, the accuracy is equal to 50, 70, 80, and 90% for rounds 1, 2, 3, and 4. Trial-based feedback will be given concerning the correctness of performance after each round (correct/incorrect). We also run control sessions that replicate our main sessions but without AI. This serves as control about learning in the task. This way, I can disentangle the impact on the observed performance of pure learning from that of AI assistance.

The third part includes the "post-believes" survey (same as the first survey) and collecting demographic information.

2.3.4 Procedures and Participants

The experiment was designed as an online task and preregistered at `https://osf.io/ub67e/`.[1] Prolific [2] was used as a platform to recruit participants and conduct an experiment. To be eligible, Participants should be fluent in English and be connected to a Desktop computer. I do not discriminate against participants on any other grounds. Participants are also asked to answer a few attention and comprehension questions. Participants were given preliminary information regarding the purpose, procedure, compensation for participation, confidentiality, and an opportunity to withdraw from the experiment.

Overall I created 43 stimuli: 3 for the training part and 40 for the main experiment. Stimuli differentiated in specularity, the number of filled elements, and rotation angle.

In total, 319 participants were recruited for the experiment. The average age was 26.4 (SD = 7.9) years old (age: M = 25.35 years, SD = 7.05; F = 28.03, SD = 8.73; 125 male, 92 female, one undefined, one missing

[1] Raw data will be stored in *.csv* format and made accessible in the same repository. The code adopted to analyse and organise the data will also be available, and the analysis will be fully replicable (RMarkdown notebook).

[2] `https://www.prolific.co/`

value).

Raw data were controlled for attentiveness with control questions. After filtering the data, I had 219 participants complete the experiment. Among them, I had 106 participants assigned to Low Time Constraints (52 nAI; 54 AI) and 113 participants - High Time Constraints (55 nAI; 58 AI). Relatively equal distribution is detected for a Penalty: High 436 (freq: 0.498) and Low 440 (freq: 0.502). The average age is 26.42, equal to the initial value of unfiltered data.

Participants received a monetary reward for participation in the experiment. Additionally, participants were given an initial endowment, and for each incorrect answer, they are fined a Low penalty (-1 point) or a High penalty (-10 points). Participants were informed in advance if the round was Low or High penalty. The final payment is given by the initial endowment minus the sum of penalties. The average payment made through the Prolific platform was 3.31 GBP.

2.4 Research Hypotheses

Based on prior research and reported earlier, the aim of this study is to evaluate the potential benefits of technology in improving human performance. Thus, the first hypothesis was formulated as follows:

H2.1: *In general, individuals perform better with an AI presence compared to a non-AI presence.*

Daily human activity involves various decision-making tasks, from short-term to long-term. Previous studies show that speed goal task leads to faster answers and less offloading, even though it may come at the expense of accuracy (Weis & Wiese, 2019a). However, having a goal to finish a task earlier and having limited time constraints are not the same issue. Therefore, I assume that with more time, individuals will be willing to recheck technology and their own results, therefore, have more chance to give a correct answer.

H2.2: *Under lower Time Constraints(lowTime-C), individuals perform better regardless of AI accuracy and penalty.*

The majority of human's daily tasks include not only accuracy and speed components but also risk component. Although scholars have been considering how humans offload cognition in tasks related to time constraints (Weis & Wiese, 2019a), it needs to be clarified how alternative time constraints affect cognitive offloading in a risky environment. Meanwhile, in a risky environment, the consequences of each mistake lead to a high penalty. This issue is particularly acute in an information technology-infused environment with increasing AI predominance. Thus, understanding the penalty influence on offloading and performance might improve the design of human-machine interactions and the complementary role of AI.

Qualitative studies showed that developers in high-risk conditions are not ready to trust technology in problem-solving situations and prefer to use mental sources (Pashchenko, Vu, & Massacci, 2020). Accordingly, I expect that since AI does not ensure full accuracy in a high-risk environment with a high penalty, participants are willing to check themselves (little offloading), which results in better performance. In contrast, in a low-risk environment with a small penalty, they will be less willing to do additional mental work and show a willingness to use an AI (to offload). Hence, to further confirm this idea under the risk component (penalty) while using AI technology, I investigate the following hypothesis:

H2.3: *Both with and without AI assistance, individuals perform better under high Penalty(highPnlt) compared to low Penalty(lowPnlt).*

Partially followed by **H** 2.2 and **H** 2.3, and going deeper into not only the presence of AI but also its accuracy, I want to confirm that AI accuracy does not have a significant impact on individuals' performance in specific conditions:

H2.4: *AI accuracy has an effect on performance: the higher accuracy, the higher performance.*

2.5 Analysis

2.5.1 Impact of AI on performance

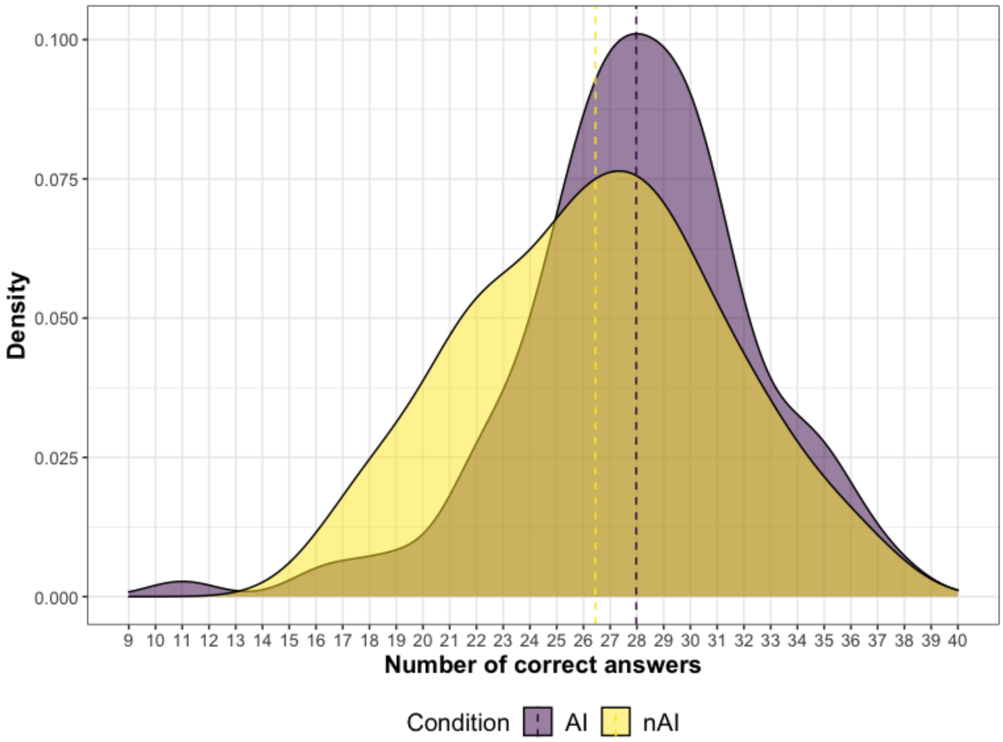
The general performance and timing of the experiment are presented in Table 2.2. Results show a difference in performance between nAI (mean 26.45) and AI (mean 27.95) sessions. To elaborate, having AI assistance, participants gave more correct answers. While the min value is smaller for the AI session, the median and max values are relatively equal and confirm the previous observation. The fact that AI influenced participants' performance also indicates that, on average, the experiment took less time for participants who were assisted by AI (22.55 min) compared to those who were performing with no assistance (24.45 min).

Fig. 2.3 presents a kernel density estimate of performance outcomes for the nAI and AI sessions. The distribution of participant performance in the nAI session is more dispersed compared to the AI sessions, in which the

Table 2.2: Descriptive statistics for nAI and AI sessions: performance and timing.

	var	mean	sd	min	q25	median	q75	max
nAI	performance	26.45	4.83	16	23	26	30	37
	time.taken (in mins)	24.45	23.03	8.09	17.58	20.44	25.77	242.15
AI	performance	27.95	4.41	11	26	28	31	38
	time.taken (in mins)	22.55	19.23	8.30	15.48	17.99	20.87	127.11

Figure 2.3: Frequency of correct answers distribution, nAI vs. AI



distribution displays higher concentration and an average is located around a higher value (*28*). These observations suggest that the incorporation of AI assistance during decision-making processes leads to performance results that are more centralised around a particular value.

Result 1. With the AI assistance participants overall performed better.

Table 2.3: Performance depending on Time constraints and AI/nAI presence.

Treatment	AI	N	Mean	SD	median
highTime-C	nAI	55	24.3	4.2	24
highTime-C	AI	58	26.3	3.9	27
lowTime-C	nAI	52	28.8	4.4	29
lowTime-C	AI	54	29.7	4.2	30

2.5.2 Impact of time constraints on performance

Time constraints also have an impact on performance. Overall, in the lowTime-C session, participants had more time to solve the task; hence, as I expected, they gave more correct answers compared to highTime-C. This variance holds both in nAI (f-test: 29.08, p-value: 4.31e-07 ***) and AI (f-test: 19.17, p-value: 2.73e-05 ***) sessions, although the nAI session reflects a slightly more significant difference. Tab. 2.3 confirms previous findings regarding performance and demonstrates that regardless of the time constraint, participants experienced a strong influence of AI assistance in quantity as well as in distribution (Fig. 2.4). As expected, in the presence of AI, participants gave more correct answers both in highTime-C and lowTime-C tasks, although in lowTime-C this difference is not significant (highTime-C: f-test: 7.32, p-value: 0.0079 **; lowTime-C: f-test: 1.352, p-value: 0.248).

Since I had four rounds with different AI accuracy on all rounds, detailed analysis by round is also essential. Fig. 2.5 provides a graphical representation of performance by time constraint and assistance in all four rounds. Confirming previous general findings, I see that in the nAI session, there is a difference in performance between lowTime-C and highTime-C tasks in all four rounds. Results of the AI session support this trend, except for the last round (f-test: 1.262, p-value: 0.264). Fig. 2.5 also demonstrates that in highTime-C sessions, participants demonstrated stable performance independently from Round when there was no AI Assistance. However, with AI

Assistance, I observe a clear trend of increasing Performance when AI Accuracy (round) increases. In the lowTime-C task, participants in both nAI and AI sessions had similar performances in round one. However, results differed starting from the second round, where the probability of correct AI suggestions increased to 70%. Moreover, in the second round (AI Accuracy 70%), performance in the AI session differs on the negative side compared to nAI, while in round 3 (AI Accuracy 80%), it differs on the positive side. This very much looks like a threshold where participants start to rely on AI suggestions rather than on their mental sources. However, in highTime-C, this assumption does not hold, and the switch is on the third round.

Figure 2.4: Performance by time constraint and assistance.

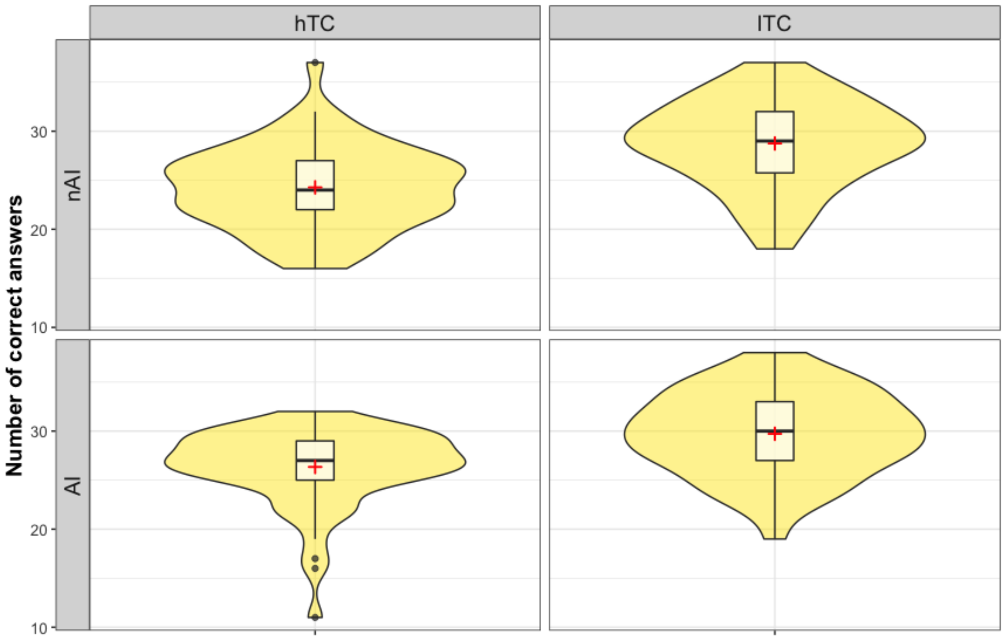
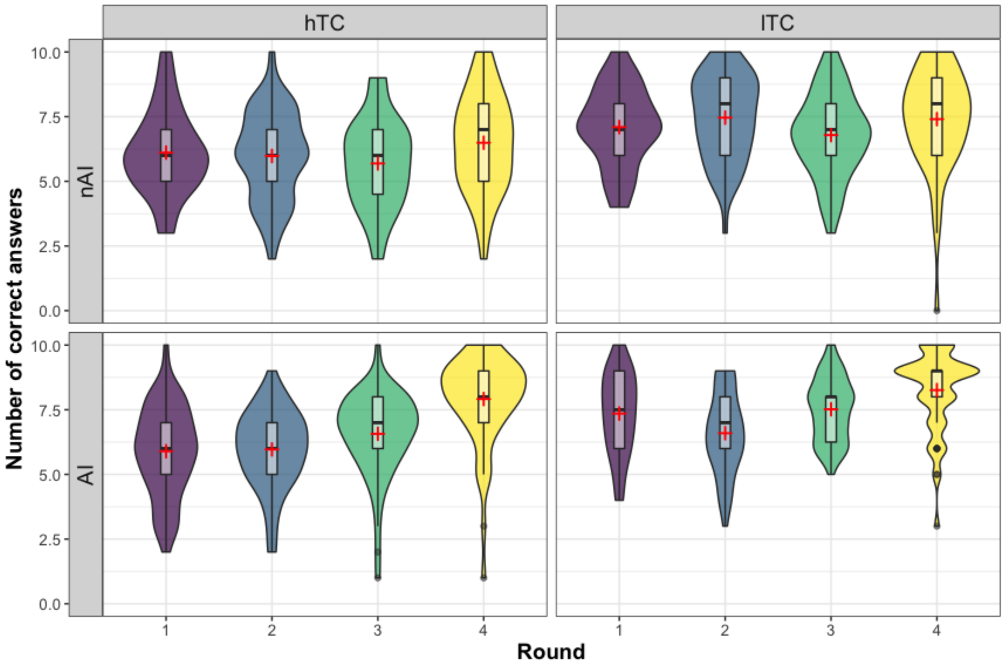


Fig. 2.6 shows the time taken to solve a task. The task took more time to complete in the nAI session (7.63 mins) compared to the AI session (5.95 mins). This trend is also true while pairwise considering time constraints jointly, thereby confirming AI influence. When AI had a relatively low accuracy (50%) in round 1, participants did not consider it

Figure 2.5: Performance by time constraint and assistance in rounds.



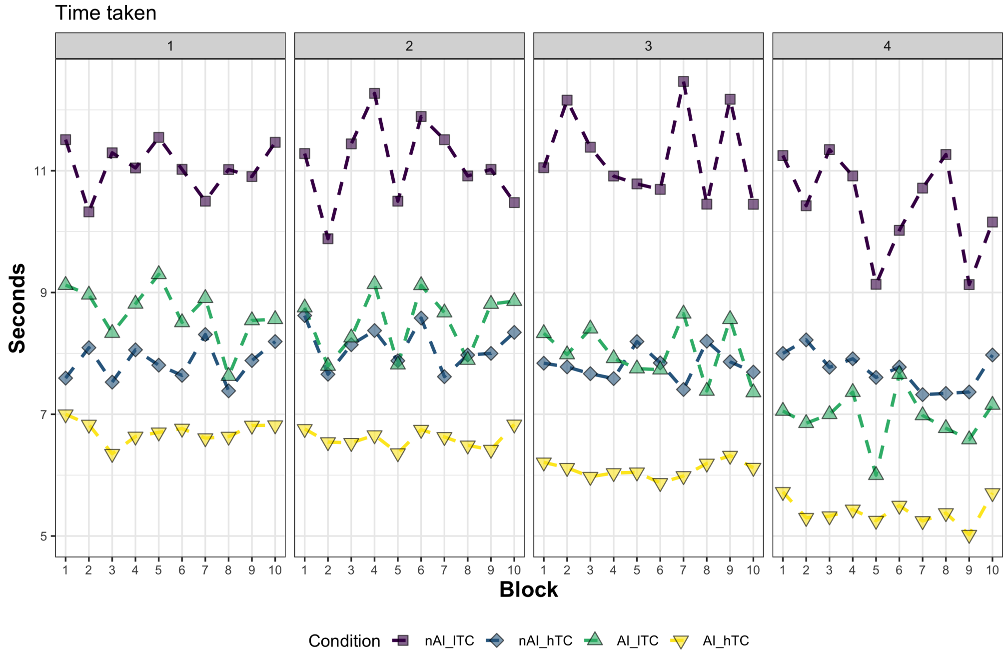
a trustworthy source. Therefore the majority of participants spent more time on both lowTime-C tasks. Although this trend persisted when the AI Accuracy was 70% (second round), with Accuracy increased to 80% in round three highTime-C in nAI and lowTime-C in AI were similar in time spent. Finally, when the AI assistant had the highest probability of correct suggestions (90%) in the fourth round, the lowTime-C time in AI dropped significantly below the highTime-C time in nAI.

Result 2. Under lower time constraints, participants spent more time on task and performed better than in high time constraints (both nAI and AI).

2.5.3 Impact of Penalty on Performance

The penalty is the other major component of our analysis. According to Table 2.4, performance was not different between nAI and AI sessions under

Figure 2.6: Timing by round and block for each treatment and session.



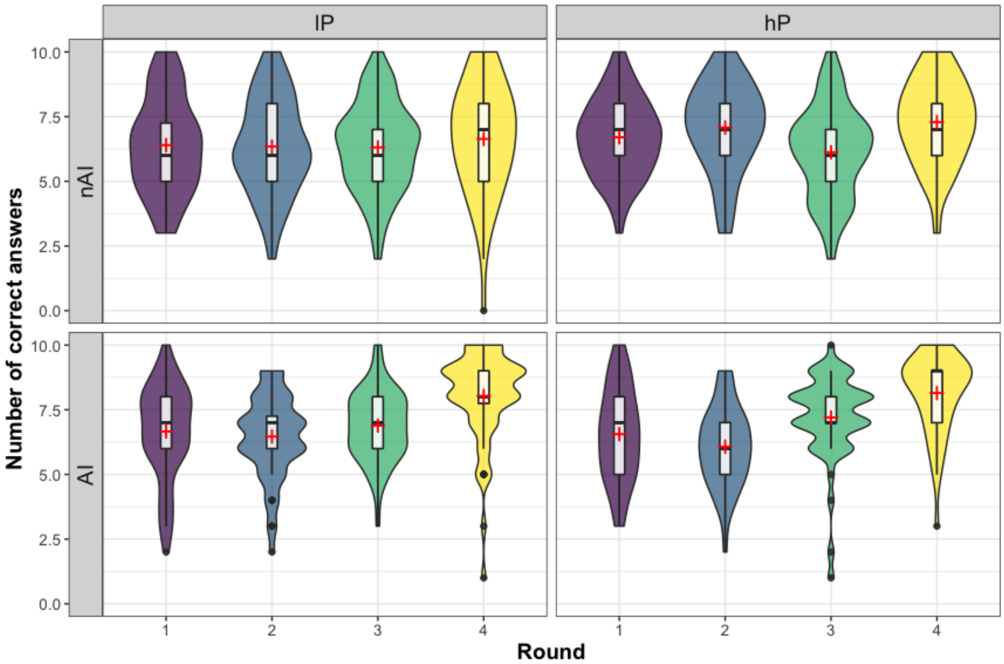
highPnlt. While comparison of means in highPnlt and lowPnlt between the two sessions suggests that without assistance, participants experienced different influence of Penalty, still in the range of standard deviation. Results in Fig. 2.7 confirm that the assigned penalty has no significant impact on performance in the nAI session, and even in the AI session, only in round two, there is a difference between Penalties (lowPnlt and highPnlt) under high time constraints (f-test: 5.483; p-value: 0.0228 *).

Table 2.4: Performance by penalty and assistance.

Penalty	AI	N	Mean	SD	median
LOW	nAI	103	13.2	6.5	12
LOW	AI	107	15.1	7.0	15
HIGH	nAI	100	14.7	6.8	14
HIGH	AI	106	14.3	6.7	13

Result 3. According to results Penalty does not have a strong impact on performance.

Figure 2.7: Performance by penalty and assistance in rounds.



2.5.4 Impact of AI Accuracy on performance

In order to investigate the influence of AI Accuracy rather than its presence and to make a more detailed and comprehensive analysis, I ran a mixed Model regression controlling for fixed and random components. A generalized linear mixed model fit by maximum likelihood (Laplace Approximation) was used to analyze the performance data. In the general equation, stepwise added interaction and control variables were to identify the best-performing model. Overall I consider four models (Tab. 2.5).

general model:

$$Performance_i \sim \beta_1 Round_i * Assistance_i + \beta_2 Treatment_i + Penalty_i + \alpha_{prolificID,i} + u_i \tag{2.1}$$

- Fixed effects: *Round, Assistance, Treatment, Penalty, Age, Gender,*

Table 2.5: Performance regression.

	(0)	(1)	(2)	(3)
(Intercept)	0.549 (0.164)***	0.378 (0.127)**	0.811 (0.335)*	0.809 (0.178)***
Round	0.146 (0.021)***	0.032 (0.030)	0.031 (0.030)	0.036 (0.030)
AI assistance	0.213 (0.065)**	-0.360 (0.123)**	-0.357 (0.122)**	-0.378 (0.140)**
Treatment, Low Time constraints	0.478 (0.065)***	0.464 (0.067)***	0.479 (0.065)***	0.570 (0.104)***
Penalty, Low Penalty	-0.098 (0.050)·	-0.103 (0.051)*	-0.098 (0.050)·	-0.118 (0.083)
Age	-0.016 (0.004)***		-0.015 (0.012)	-0.016 (0.004)***
Gender, male	0.152 (0.068)*	0.203 (0.069)**	0.152 (0.070)*	0.152 (0.067)*
Gender, other	0.091 (0.357)	0.183 (0.370)	0.091 (0.359)	0.111 (0.359)
Risk attitude	-0.014 (0.014)	-0.011 (0.014)	-0.010 (0.055)	-0.014 (0.014)
Round:AI assistance		0.234 (0.042)***	0.234 (0.042)***	0.225 (0.043)***
Age:Risk attitude			-0.000 (0.002)	
AI assistance:Treatment, Low Time constraints				-0.065 (0.129)
AI assistance:Penalty, Low Penalty				0.139 (0.101)
Treatment, Low Time constraints:Penalty, Low Penalty				-0.111 (0.101)
AIC	10739.241	10722.431	10712.472	10713.245
BIC	10810.020	10793.210	10797.407	10812.337
Log Likelihood	-5359.620	-5351.215	-5344.236	-5342.623
Num. obs.	8760	8760	8760	8760
Num. groups: prolificID	219	219	219	219
Var: prolificID (Intercept)	0.103	0.119	0.104	0.100

p <0.001; p <0.01; p <0.05; ·p <0.1

Risk attitude;

- Random effects: *prolificID*;
- Interaction effects: *Round:Assistance*;

model (*p0*): In the first model, I employed all basic variables like Round, Assistance, Time constraint, Penalty, and also integrated additional variables like Age, Gender, and Risk attitude. The findings indicate that Round and lowTime-C (Time constraint) have a significant positive effect on performance. In other words, as the round increases, individuals give more correct answers, the same as when they have more time to solve a task. Furthermore, an overall positive effect on performance was observed in relation to the presence of AI.

model (*p1*): We estimate a model with an interaction effect of Round and Assistance. Results show that the interaction effect is positive and significant. However, separately these components have opposite effects. Round is not significant, which confirms that there is no learning effect across the rounds. Moreover, AI Assistance is significant and decreases performance.

model (*p2*): The following model is similar to the previous one, except I included variance Age. Results show that as from the previous regression, AI Assistance, lowTime-C and interaction effect of Round and Assistance are significant. I also see that, unlike the previous model, when age is considered, the impact of the Penalty becomes weaker. Age itself has a negative and substantial impact on performance.

model (*p3*): The last model adds all possible interaction effects between three variables: Assistance, Time Constraint, and Penalty. The main results are confirmed, but as shown by the information criteria, the additional explanatory variables do not improve the explanatory power of the model. Therefore model can not be considered as good.

Conclusion: Given the information criteria the model ***p2*** was chosen as the best regression model.

As I see from Table 2.5 coefficient of AI assistant is negative (p-value <0.01), which means that the presence of AI itself has a negative effect on performance. Farther regression confirms the significance of time constraint, showing that under lowTime-C, participants having more time performed relatively better. Regression also demonstrates that the interaction effect between Round and AI Assistance has a positive coefficient (significant). This interaction effect shows that when AI was present, and its accuracy increased (round by round), Performance also increased. The coefficient of Low Penalty in Table 2.5 has a negative sign, which means that the higher the Penalty, the better Performance.

Analysis of regression and Fig. 3.5 allow us to say that both with and without AI assistance, individuals perform better under high Penalty(highPnlt) compared to low Penalty (lowPnlt). However, the only significant difference between sessions was in round 2 (with 70% of AI accuracy) for the AI session.

Result 4. AI accuracy has an effect on performance.

Table 2.6: Reaction time regression.

	(t0)	(t1)	(t2)
(Intercept)	**8.716 (0.524)*** **	**8.182 (0.529)*** **	**8.110 (0.551)*** **
Round	**-0.262 (0.023)*** **	-0.054 (0.036)	-0.064 (0.036)·
as_factor(assist)AI	**-1.681 (0.229)*** **	**-0.814 (0.257)** **	**-0.752 (0.337)* **
as_factor(Treatment)lowTime-C	-0.018 (0.228)	-0.027 (0.228)	0.128 (0.332)
as_factor(Penalty)lowPnlt	-0.084 (0.058)	-0.102 (0.057)·	0.114 (0.101)
as.numeric(age)	0.008 (0.015)	**0.008 (0.015)**	**0.007 (0.015)**
as_factor(gender)female	-0.185 (0.238)	-0.180 (0.238)	-0.165 (0.239)
as_factor(gender)other	-2.070 (1.180)·	-2.068 (1.179)·	-2.083 (1.187)·
as.numeric(Risk_attitude)	0.000 (0.049)	0.003 (0.049)	**0.004 (0.049)**
Round:as_factor(assist)AI		**-0.338 (0.046)*** **	**-0.327 (0.046)*** **
as_factor(assist)AI:as_factor(Treatment)lowTime-C			-0.026 (0.457)
as_factor(assist)AI:as_factor(Penalty)lowPnlt			-0.166 (0.119)
as_factor(Treatment)lowTime-C:as_factor(Penalty)lowPnlt			**-0.279 (0.117)* **
AIC	22527.443	22478.989	22481.095
BIC	22600.082	22558.231	22580.149
Log Likelihood	-11252.722	-11227.494	-11225.548
Num. obs.	5451	5451	5451
Num. groups: prolificID	219	219	219
Var: prolificID (Intercept)	2.587	2.584	2.595
Var: Residual	3.225	3.192	3.188

$p < 0.001$; $p < 0.01$; $p < 0.05$; $\cdot p < 0.1$

We ran also regression for reaction time (Table 2.6).

general model

$$ReactionTime = Round{*}Assistance + (1|prolificID) + Treatment + Penalty \tag{2.2}$$

model (*t0*): In the first model, I used the same variables as for model p0. I see that the only significant variables are Round and Assistance.

model (*t1*): The second model was enriched by the variable interaction effect of Round and Assistance, which is significant. However, in this model, Round is not significant anymore. I also note that the variable Age is significant and has an effect on Reaction time.

model (*t2*): The last model, as for p3, adds all possible interaction effects between three variables: Assistance, Time Constraint, and Penalty. Although in this model, I have a significant Risk attitude and the interaction effect between Time constraint and Penalty, I can not confirm that this is the best model.

The round coefficient (see table 2.6) might be misinterpreted as a learning effect in model ***t0***, as the interaction with AI is missing. Actually, with the introduction of the interaction term in ***t1*** and ***t2***, the coefficient Round becomes not significant. These regressions further confirm the effect of age on timing. Therefore age should be considered an important variable in the analysis of technology-human interaction in future studies.

Conclusion: Given the information criteria the model ***t1*** was chosen as the best regression model.

Regression results suggest that AI assistance and its accuracy significantly reduce the time participants spend performing a task. At the same time, Age might increase the length of task performance time.

Overall, results suggest considerable economic significance, as they provide insight into the factors that affect human performance when interacting with AI assistants in a risky environment. Results suggest that the presence of AI has a negative effect on individuals' performance. This suggests that the presence of AI in itself may actually impair the performance, rather than enhance it if no proper support on training is provided (on individual or organisational levels). Thus, Hypotheses 2.1 can not be confirmed. At the same time, the accuracy of the AI is crucial - the significant interaction effect between round (accuracy increased over rounds) and AI assistance indicates that AI can have a positive effect on performance if accuracy is improved over time. Thus the ability to learn and improve is a crucial aspect when designing and selecting AI assistants, both at the individual and organisational levels in order to improve performance outcomes.

Additionally, both time constraints and the penalty influence people's choices and performance consequently, partially confirming 2.2 saying that under lower Time Constraints (lowTime-C), individuals perform better regardless of AI accuracy and penalty, and 2.3 - both with and without AI assistance, individuals perform better under high Penalty (highPnlt) comparing to low Penalty (lowPnlt).

2.5.5 Survey

Before and after the experiment, I conducted a Survey to test pre/post-self-reported beliefs towards technology and AI (Fig. A.1). Surprisingly participants show no change regarding trust before and after the interaction with the financial risk and AI A.1. However, at the same time, more people agree that AI only "supports" their decision. Therefore, we can conclude that interaction with AI does not affect trust. However, it increases human confidence having a significant role in decision-making.

Predictably, after experiencing time manipulation tasks, participants, those who interact with AI, experience increased determinacy concerning the question of AI assistance/interference under time constraints. Additionally, most agreed that an AI could solve precision/computational tasks better than humans. At the same time, participants were relatively uncertain regarding AI superiority in Problem-solving tasks before the experiment (both with nAI and AI sessions), while after, their positive attitude increased.

In technology-human interaction, the majority tend to have a positive attitude towards technology when they first use it until it gives a reason not to. However, after the experiment, we noticed a few contrary results: those who did not interact with AI increased positive attitudes. At the same time, those who had AI assistance slightly decreased their assurance in AI. Finally, the majority is more confident and confirm that they will continue to use smart assistance and information technologies, especially after the experiment, even if it fails them several times.

I also calculated Cronbach's alpha to report measures of internal consistency and possibly include it in the regression. However, results show that the survey is unreliable since the index is equal to 0.43 (0.36)[3]. Nonetheless, it should not rule out the survey's insight for overall analysis.

[3]for the first and the second surveys, respectively

2.6 Discussion

Although participants generally performed better under AI assistance, results demonstrate that the AI presence negatively impacts overall performance. This inconsistency happens because of somewhat scattered probabilities of correct suggestions of AI. For example, in the first rounds, suggestions were quite unreliable, while AI gave 9 out of 10 correct suggestions in the last round. Thus, blindly following AI suggestions, one can obtain a very high score in the last round but fail dramatically in the first. We conclude that participants offload cognitive tasks on AI even if it is not accurate and safe. The narrow concentration of performance results around a specific value also demonstrated the tendency to offload cognitive tasks on AI. When participants had no AI support, results were much more varied, which further illustrates that human beings can easily blindly follow AI's guidance. The issue of trust in AI depends on its accuracy. Findings suggest that participants having trustworthy AI assistance will follow it, regardless of penalty and time constraints.

Participants also spent more time on tasks under lower time constraints and performed better than under high time constraints (nAI and AI). Additionally, it was found that under lowTime-C, the difference in performance could be more evident. The latter could be explained by the fact that under lowTime-C, participants had enough time to solve the problem using internal resources. It was also found a threshold of reliance on AI. Furthermore, when participants have less time to solve a task threshold of reliance on AI switch to higher values compared to lowTime-C. A probable explanation for this is that participants in highTime-C had less time and focused more on the task, and the threshold for reliance on AI should be higher.

The Penalty was a significant component of our analysis which should have a strong influence on performance. However, results show that although Penalty affects performance, it needs to be stronger. The Penalty was not high enough to make an impact on a performance given already treatments like Time constraints and AI assistance. The small difference in the Penalty values caused the effect of the Penalty to be so small that it became insignificant. Therefore, in future studies, Penalty should be adjusted to a greater variety or/and assigned separately from the valuable treatments. This might be tested in future studies.

A complementary finding in our study is age's influence on performance and reaction time. I found a small but significant impact on reaction time - younger participants performed the task more quickly. Age has a negative impact on performance, which might be nothing in common with automation. Still, this variable should also be taken into account in future studies.

The results of the paper help to understand how speed and accuracy goals affect cognitive offloading in a risky environment. This is an essential element in designing human-AI synergy. The main objective of RQ is to help to identify critical parameters that influence a human decision to use external resources instead of brain-based internal resources in a risky environment. Understanding vulnerable layers, where humans might wrongly offload or decide to solve the problem on their own, will allow for correctly curing and designing those parts, which in turn, will save a considerable amount of fines incurred due to incorrect decisions. Therefore, the study will help to save investments in designing human-AI systems and increase the understanding of human-technology interaction, ultimately contributing to a healthier and more thrifty use of resources.

Integrating AI technologies with behavioural science can help solve a wide range of policy and social problems. Collaboration is vital, and designing behavioural interventions with AI solutions has the potential to encourage "healthier" behaviour without restricting choice but with a considerable impact on the economy. An excellent balance and the ability to look at AI as a component in decision-making that is as important as the human viewpoint is the future of financial decision-making. Utilizing AI to help individuals make more optimal financial decisions comes with both individual and societal benefits.

2.6.1 Limitations

The experiment design for some could be counter-intuitive as it seems that human problem-solvers can easily use AI help to re-check their decisions. However, the goal of the experiment is not to discover that humans use AI help but to check in which exact situations, depending on the performance goal, they will use AI, and if so if there might be a habituation effect. The second limitation concerns our technical issue related to the nAI session. During the experiment, some participants reported a short server interruption, which was restored for the next task. This could affect the participants' answers in the sense that they could lose points for an unanswered question because the absence of an answer was counted as an incorrect answer. However, our analysis confirms that this issue did not affect the data acquired.

Supplementary materials can be found in *Appendix A*.

2.7 Conclusions

The present study aimed to investigate the impact of the penalty on offloading behaviour when individuals interact with AI technologies, specifically how performance is affected when individuals are confronted with different levels of external pressure, as measured by the time available to complete a visual task. The study sought to establish and clarify the relationship between human decision-making and possible cognitive offloading on an algorithm, depending on the environment in which an individual is constrained in deliberation time. In the study, participants performed better with AI assistance, but the presence of AI negatively impacted overall performance. The scattered probabilities of correct AI suggestions caused inconsistency in performance results, demonstrating cognitive offloading on AI, even if the suggestions were unreliable. Therefore, it is crucial to understand the factors influencing and circumstances in which people are more likely to offload critical decision making to an algorithm, potentially to their detriment. As an example, in our study time constraints had a significant impact on performance, with participants performing better under lower time constraints, thus participants were likely to rely more on AI when they had less time to solve the task. At the same time, penalty had a minimal impact on performance in the presence of AI assistance, which probably should be stronger and assigned separately from time pressure in future studies to investigate its true impact. Understanding circumstances and factors influencing reliance on algorithms can be used to prevent erroneous actions on the part of users of hybrid intelligent systems, facilitate the decision-making of over-confident users, and provide support to users vulnerable to the influence of AI and algorithms.

Since partnering with machines and smart assistants is becoming almost inevitable, an additional important topic arising from chapters one and two is individual financial decision making and the aspects influencing it. One of the big problems and bias is a tendency to rely even more on the outcome or result rather than on the strategies of decision making or intentions. In the next chapter, I will expand on the previous topic and discuss how the information about the action's outcome affects the human incentive behaviour depending on the interacting partner and how the presence of the non-human partner will affect the reward and punishment schemes. The aim would be to confirm the presence of outcome bias and the influence of a random outcome of the risky investment on the Principal's evaluations and rewards for the Agent or refute.

Chapter 3

Ain't blaming you: Delegation of financial decisions to humans and algorithms

3.1 Introduction

People tend to evaluate decisions after the fact, meaning that evaluation of the quality of the decision made includes information not only about decision strategy but also the result of the outcome. Moreover, in real-life events, too many factors need to be controlled. Therefore, the quality of the decision made is not necessarily related to the quality of the outcome. Therefore, humans tend to rely more on the outcome or result rather than on the strategy of the decision. For instance, Steven Bradbury won the Olympic gold medal in short-track speed skating in 2002, not the fastest but the only one who reached the finish line. Bradbury's story is an excellent example of how the "right" practice strategy of trying to be the fastest failed, and the losing strategy of trailing at the end of the race led

Reported in this chapter is the result of a joint work with Matteo Ploner.

to winning the gold medal and becoming a national hero. This example is not about which strategy is right and which is wrong; rather, it gives us an example of how human beings might make inferences from the outcome ignoring the tactic and strategy. The tendency to make inferences from the outcome about the strategy quite often might lead to a situation when humans confirmed the chosen strategy as the right decision when the outcome was favourable (Lipshitz, 1989), even if the decision was wrong, but the outcome was affected by a random event or by chance. A phenomenon when people tend to rely more on information about the outcome when judging decision quality rather than the decision-making process is known as *outcome bias* (Baron & Hershey, 1988).

From more trivial examples, when we have an individual, her actions and the result of her actions, an increasing amount of interactions between an individual and artificial agents take place in the real world. The statement that intelligent algorithms are increasingly interfering with human lives is a truism. The further the digital society develops, the more anchor points there will be for automating processes in social, economic and even cultural spheres. Technology, as humans' "secret assistants", could perform a variety of tasks, including non-computational nature tasks and perform actions that require the processing of meaningful information and are considered the prerogative of the human brain. Among more implicit roles, for example, scholars suggest that AI advisors are put in the role of a scapegoat to which one can deflect (some of the) moral blame for dishonesty (Leib, Köbis, Rilke, Hagens, & Irlenbusch, 2021). At the same time, for example in a financial market setting humans may be prone to over-relying on autonomous robots, leading to poor trading outcomes (Asparouhova et al., 2020).

In this essay, I examine how the information about the result (outcome) of the investment affects the reward and punishment behaviour of the participants that interact with Human and Algorithm agents. Specifically, I conduct an experiment investigating the interaction between outcome bias and human/algorithm responsibility. Previous experimental evidence has shown that this bias is widespread in investment evaluations in the laboratory (König-Kersting, Pollmann, Potters, & Trautmann, 2021), also in experiments with finance professionals, scholars found that humans' own decision-making quality is associated with a decrease in delegation frequency (Holzmeister, Holmén, Kirchler, Stefan, & Wengström, 2022). Additionally, scholars demonstrate that humans generally relied more heavily on algorithms when the market was complex and decision-making was difficult (Asparouhova et al., 2022). Therefore, this paper will test whether outcome bias is more pronounced when investment choices are made by humans or by an algorithm and whether the delegation decision depends on the information available to the Principal.

I designed an experiment by manipulating information available: Ex-ante, Ex-post and participants' possibility of task delegation: no delegation or algorithm delegation. In the experiment, the Principal delegated the Agent a risky financial decision. In the experiment, the participant with an initial endowment (Principal) delegated a risky financial decision to another participant (Agent). The Agent could solve the investment task on her own or delegate the investment task to the algorithm. The Principal, in turn, reward or punish the Agent for her choice. Results confirm the presence of outcome bias and the influence of a random outcome of the risky investment on the Principal's evaluations and rewards for the Agent. I also found that the outcome bias does not differ when a choice is made personally or delegated to the algorithm. Finally, delegation to an algorithm and the level of risky investments do not differ when knowing that rewards or punishments are given before or after the investment outcome.

The current experimental study adds to several strands of related research. First of all, we contribute to the growing literature on outsourcing investment decisions to automated algorithms and robo-advisors. This topic has gained significant attention from the scientific community (Rossi & Utkus, 2020; D'Acunto, Prabhala, & Rossi, 2019). Previous studies found the presence of outcome bias in investors' behaviour (Germann & Weber, 2018) and in principals' evaluation and rewards for financial agents in risky investment decisions (König-Kersting et al., 2021). Our study further confirms outcome bias presence in the field of investment decisions and principals' tendency to evaluate the agents' performance based on the outcome.

Second, we add to the literature on interaction partners in delegation investment decisions. We demonstrate that the outcome bias does not depend on the source of the investment choice. Thus outcome bias is not mediated by delegation to an algorithm. Previous research found that principals judge the same decision differently despite their exact knowledge of the investment strategy (König-Kersting et al., 2021). In our study, we go further by adding an artificial agent to the interaction.

Furthermore, this work contributes to the literature on shifting blame. In the experiment, the agent does not delegate more to the algorithm in the Ex-post condition than in the Ex-ante condition. Thus, the agents do not generally "hide" behind the machine, which may be interpreted as no shift in responsibility to save on self-image concerns.

Our findings may have implications for growing financial markets for several reasons. We demonstrate the potential pathways which are important to consider when designing human-algorithm interaction. First of all, as our study shows, we still have to consider the presence of outcome bias in the investment field working with artificial agents and work on solutions to minimise the bias as much as possible. Additionally, the current study demonstrates that individuals in a high-risk environment do not exhibit a distinct preference for using algorithms more than other humans to avoid or deflect responsibility and shift the blame. This suggests that in developing a framework for human-machine interactions human can equally entrust both agents with responsibility and not use algorithm agents as a scapegoat. Finally, despite the rise of robo-advising, recipients of the decision making and responsibility attribution will still be held by a human. And systematical analysis of humam-algorithm interactions can enhance our comprehension of assigning responsibility to respective parties.

3.1.1 General framework and motivation.

Intentions are defined as elements of partial plans of activity to which the human has committed themselves (Bratman, 1987). Saying so, a human having an intention could be represented as a human with internal states with representational content originating him to a subsequent action. Though, not all intentions are summarized in actions. Often humans fail to implement their intentions into actions and achieve the goal, known as an Intention Action Gap. In other words, this gap happens when humans' values, attitudes, or intentions do not match their actions. On an individual level, this could happen due to bias towards immediate gratification or inaction caused by setting up an elusive goal. Moreover, speaking about the intentions and actions of other external subjects, a human cannot generally observe the link between their intentions, actions or inactions. What could be possibly observed, however, is the consequences of the subject's actions.

There are different combinations of observed actions and consequences, and depending on the combinations, they might be relevant for different fields. For example, for managed funds, it is mainly observed consequences or outcomes but not actions, while for credence goods, it is mostly observed actions but not consequences.

According to Baron (Baron, 2000): "Rationality concerns the methods of thinking I use, not the conclusions of our thinking. Rational methods are generally best in achieving the thinker's goals." Outcomes should matter only in conjunction with strategies, especially when uncertainty is present. However, a human being is often detached from rationality.

Decision evaluations occur after the fact and therefore include information about outcomes (consequences), even though this is often not the best indicator of decision quality (Lipshitz, 1989). Though outcome bias is widespread and manifests itself in various domains, from human resource management decision-making (Bankins, Formosa, Griep, & Richards, 2022) to sports (Lefgren, Platt, & Price, 2015). For example, in ethical decision making, Gino et al. (Gino, Moore, & Bazerman, 2009) showed that the same behaviour generates more ethical condemnation when it results in bad rather than good outcomes, even if the outcomes are determined by chance. Unpleasant discovery of outcome bias in the medical field demonstrates that information about outcomes might affect the overestimating of the quality of the medical decisions, the difficulty of making a diagnosis, and memory of weights assigned to signs and symptoms consistent with diagnosis (Sacchi & Cherubini, 2004; Caplan, Posner, & Cheney, 1991), which could have an adverse effect on doctors education and knowledge testing processes.

Given that outcome, bias tends to increase when people have little or limited information to evaluate the quality of a decision (Baron & Hershey, 1988). The financial sector, with its uncertainty, seems to be an incredibly challenging domain for an average person. Indeed, investors show clear tendencies towards outcome bias in financial decision making (Germann & Weber, 2018). From a human being's perspective, it is often assumed that if an investment strategy was successful or had a positive outcome, the strategy should be rated as more valid and result from better decision-making.

Furthermore, it is difficult to prove to a decision maker that this inference is not necessarily correct and that they should avoid falling into outcome bias. Moreover, even if people have been told to ignore the outcome bias when judging decisions, they failed to do so (Fischhoff, 1975). Correctly recalling the probability of the outcome also did not help participants to avoid the influence of emotional responses leading individuals to switch from more profitable lotteries to less profitable lotteries after observing unfavourable outcomes (Ratner & Herbst, 2005).

However, it is unclear whether outcome bias might be shifted or diminished if humans perform in the financial field with an algorithmic partner and whether a non-human partner might influence the outcome-strategy relation.

3.1.2 Do choices made by a machine shift attention from the outcome to the process?

The statement that intelligent algorithms increasingly interfere with human lives is a truism. The further the digital society develops, the more anchor points there will be for automating processes in social, economic and even cultural spheres. As humans' "secret assistants", technology could perform various tasks, starting from financial tasks like being an advisor (D'Acunto & Rossi, 2021), to non-computational nature tasks and perform actions that require the processing of meaningful information and are considered the prerogative of the human brain. Among more implicit roles, for example, scholars suggest that AI advisors are put in the role of a scapegoat to which one can deflect (some of the) moral blame for dishonesty (Leib et al., 2021).

In terms of human-algorithm interaction, it is found that outcome bias applies not only to human decision-makers. For example, evidence of outcome bias application to AI decisions in human resource management decision-making was demonstrated by Bankins et al. (Bankins et al., 2022). Additionally, in the medical field, scholars found that for resource allocation decisions, there is evidence of an outcome bias for both human and AI decision makers (Formosa, Rogers, Griep, Bankins, & Richards, 2022).

While there are findings of outcome bias in human algorithm interactions in the financial field, this topic should be studied more. Pelster and Breitmayer (Pelster & Breitmayer, 2019) showed that outcome bias is relevant in real-world financial decision making. They found that on a social-trading platform, the likelihood of obtaining likes or copiers increases in the other trader's past performance and is not affected by risk-taking. Holzmeister et al. (Holzmeister et al., 2022) had a finance professional in their experiment and delegated to the other human; what if the choice would be to delegate only to the algorithm or perform the task himself?

Koenig-Kersting et al. (König-Kersting et al., 2021) investigate the outcome bias in experimental financial decisions. In the experiment, a Principal could reward an Agent who made a choice in an Investment Task either *before* knowing the outcome or *after* knowing it. Scholars found that rewards in *after* are generally higher for successful investments than for unsuccessful investments, while rewards in *before* are in between. In other words, the principal's evaluations and financial rewards for the Agent are strongly affected by the random outcome of the risky investment. This attitude goes under the label of outcome bias as it values events that are out of individual control (random outcomes) and disregards events under one's control (intentions).

Thus, previous research found humans more likely to delegate to the algorithm than to finance professionals (Holzmeister et al., 2022). However, the evidence is mixed. Algorithms being perceived as more efficient, cheaper, and less biased than human advisors, also perceived as being less trustworthy, less personal, and less empathetic (Brenner & Meyll, 2020).

With the growth of technological assistance, and in particular algorithmic investment funds (Harvey, Rattray, Sinclair, & Van Hemert, 2017) and algorithm advisory (D'Acunto et al., 2019), it still needs to be better explored how technological assistants as decision-makers affect the choice of strategy and hybrid decision making. The intent is to investigate how this human-algorithm symbiosis would affect outcome bias and whether it shifts the attention from the outcome to the process. In our experiment, I further examine the presence of outcome bias in evaluating investment decisions.

3.1.3 Shifting responsibility and delegation of decisions.

People quite often offload and delegate tasks to algorithms, especially computational tasks (Osiurak et al., 2018; Walsh & Anderson, 2009). However, in some cases, they do not rush into the aid of algorithms. A human might avoid interaction with algorithms because the task might be subjective in nature or moral questions, and people tend to see that algorithm is not suitable for it. This case in which a human tries to avoid the advice of an algorithm or, in general, to avoid interaction with the algorithm is called algorithm aversion (Castelo et al., 2019). For example, in the medical field, human resistance to medical AI is caused by not considering the unique characteristics and individual approach (Longoni, Bonezzi, & Morewedge, 2019). In the financial field, humans exhibit algorithm aversion when they perceive the algorithms to be opaque, complex, not transparent, or have low accuracy (Germann & Merkle, 2022; Mahmud, Islam, Ahmed, & Smolander, 2022), or lack experience with algorithmic decision making (Filiz, Judek, Lorenz, & Spiwoks, 2021). Additionally, mistakes made by the algorithm turn people away from them and make them less confident in it, more than mistakes made by another person (Dietvorst, Simmons,

& Massey, 2015). Humans are inclined to be less confident in algorithms (Harvey et al., 2017), moreover mistakes made by the algorithm turn people away from them and make them less confident in it, more than mistakes made by another person (Dietvorst et al., 2015). Presumably, algorithm avoidance might cause a slowdown in intelligent technologies implementation (Niszczota & Kaszás, 2020).

Yet, given the many benefits of smart technologies implementations, beneficial would be to find a way to eliminate these obstacles. Especially considering that evidence on algorithm avoidance is mixed (Germann & Merkle, 2022; Logg et al., 2019; Chugunova & Sele, 2022). For example, increasing human likeness increases the use of algorithms. People are also fine using algorithms after they (even slightly) modify them (Dietvorst, Simmons, & Massey, 2018). Therefore the key is to find the optimal conditions and modalities to integrate algorithms, as for example Capponi et al., (Capponi, Olafsson, & Zariphopoulou, 2022) suggested personalized robo-advising (PRA) framework combining machine learning, behavioural economics, and optimization techniques to improve investment outcomes for clients.

One of the hidden but significant issues that can cause the intention to delegate the task is a desire to avoid blame and responsibility. Decision-making and the outcome of the decision made obviously entail such concepts as responsibility and blame attribution. Indeed, the issue of the distribution of responsibility for making a decision and attributing responsibility for the actions taken is a very hot topic for discussion these days and is particularly relevant concerning algorithms. Tools and ways to deflect responsibility vary. Delegation is one of the tools to avoid responsibility for decision-making and probably subsequent punishment. For example, scholars have shown that people deflect responsibility for their behaviour to advisors (Bonaccio & Dalal, 2006), deflect blame for negative outcomes on other people (Bazerman & Gino, 2012), and also on algorithms (Hohenstein & Jung, 2020).

Responsibility attribution can be effectively shifted and constitute a strong motive for the delegation of a decision making (Bartling & Fischbacher, 2012). Bartling and Fischbacher (Bartling & Fischbacher, 2012) investigate delegation of responsibility in an allocation game involving three parties. They found that a Principal can delegate an Agent to make an allocation choice that also affects the third party. As a result: when the third party is damaged, he/she punishes the Agent but not the principal, and secondly, delegation is a way to shift the blame.

It is also true when the delegee is powerless (Hill, 2015) or if the delegation decision itself eliminated the possibility of a fair outcome (Oexl & Grossman, 2013). Going further, Feier et al., (Feier, Gogoll, & Uhl, 2022) sought to find if this also holds for artificial agents. They found that, indeed, for decision makers, delegation to an algorithm is even more effective than delegation to another human in avoiding punishment. Dana et al. (Dana, Weber, & Kuang, 2007) further show that individuals may delegate allocation choices to an algorithm to preserve self-image. Therefore findings suggest that the nature of the Agent affects the delegation decision itself and the reward and punishment that the delegator receives (Feier et al., 2022).

In their experiment, Casal et al. (Casal, Ploner, & Sproten, 2019) allowed a Principal to communicate the desired investment level to an Agent. They found that agents are incentivised to "over-invest" but can get punished by Principals. Secondly, deviations from the desired investment level get punished *only* when the investment is not successful.

Intention. Blame attribution also depends on intention because assigning intention to someone's actions leads to greater blame than others, who believed, had no intention given that both had the same consequences.

(Ames & Fiske, 2013; Hidalgo, Orghian, Canals, De Almeida, & Martin, 2021). Intentional acts are seen as objectively more harmful than unintentional ones, even when the outcomes are identical. Concerning technological interaction, people might delegate their actions to the algorithms believing that the algorithm has no intention and even in the worst-case scenario, neither the algorithm nor the human who delegates the decision-making to it will not be punished hard. Partially supporting this idea, Furlough et al. (Furlough, Stokes, & Gillan, 2021) shows that people, in general, tend to blame other people more than robots; however, the degree of blaming a robot increases to almost the same level as a person if the robot is considered autonomous. Therefore, there is an idea that a human might anticipate and expect that in conditions where the information about the outcome is available, the judgment about their action would be harsher. To put it another way, if a human expects punishment for a riskier outcome - she will prefer to delegate more to the algorithm. If a human expects less risk and punishment, she might prefer to take responsibility.

3.1.4 Risk-taking, if a human makes a decision on her own, would she act riskier or safer?

Decision-making is a multi-component process that, among others, includes taking actions and reviewing the consequences of the actions. Taking action and the consequences for actions taken can often be very attractive to push onto someone because of the foreseen responsibility and blame for a bad or unpleasant outcome. If a human decides to take responsibility and action on her own, would she act more courageous or safer? Moreover, would the risk level depend on the impending judgment from the other Human (principal)?

The question of our risk preferences changing when humans make decisions for others has yet to be answered clearly and consistently. There were no overall self-other differences in the financial domain; however, there was a moderating effect of frame: decisions in a gain frame were more risk-averse for self than other, whereas decisions in a loss frame were more risk-seeking for self than other (Batteux, Ferguson, & Tunney, 2019). If Human keeps control - it assumes a riskier investment because less info means less punishment. Alternatively, she will make a safer investment because more info is available and more punishment is anticipated.

Based on all the above-discussed, I plan to formulate our hypothesis below in the next section.

3.2 Research Hypotheses

Proceeding from the above-described, I test the following main hypotheses:

H3.1: [Outcome bias]: *For a given investment level, higher (lower) rewards are given to Agents in* ***After*** *relative to* ***Before*** *when a positive (negative) outcome is registered.*

This hypothesis follows from the assumption that the Principal attaches a positive value to the outcomes, given a choice made by the Agent. Thus, when the outcome is positive, a positive extra reward is given to the Principal. In contrast, when the outcome is negative, an extra negative reward is given to the Principal relative to the condition in which only intentions are known.

To test this hypothesis, I will compare the rewards of Principals in *Before* (Ex-ante) and *After* (Ex-post) for a given investment level and different outcomes, controlling for delegation choices.

H3.2: [Source]: *The outcome bias depends on the source of the investment choice.*

The assumption is that outcome bias positively interacts with Agent's responsibility in the choice. Specifically, when a positive outcome is observed, a higher reward is expected for an Agent who retains the choice than for an Agent who delegates to the algorithm. In contrast, when a negative outcome is observed, a stronger punishment is expected for an Agent who retains the choice than for an Agent who delegates to the algorithm. At the same time, I expect the difference in rewards conditional upon Agent's choice to be smaller in Ex-ante than in Ex-post. This will make the outcome bias larger for choices not delegated to the algorithm.

To test this hypothesis, I will compare rewards in Ex-ante and Ex-post for a given investment level, for different outcomes, and different delegation choices of Agents.

H3.3: [Delegation]: *Agents delegate more to the algorithm in the After condition than in the Before condition.*

The Agents will anticipate that Principals will evaluate negative outcomes negatively in After (Ex-post), irrespective of investment choice, and to avoid punishment, they will alleviate their responsibility by delegating the algorithm. On the other hand, in the Before (Ex-ante) condition, Agents will retain the choice to seek rewards.

To test this hypothesis, I will compare delegation choices by Agent in Ex-ante and Ex-post.

H3.4: [Risk taking]: *Agents who retain control will make riskier in-*

vestments in Before condition than in After condition.

The Agents anticipate that Principals evaluate the investments with higher expected returns positively but evaluate bad outcomes negatively. Thus, they will seek rewards when the outcomes are not known (Ex-ante) by taking higher risks without the threat of being punished for the negative outcome.

To test this hypothesis, I will compare the investment choices of Agents who retain control over the investment in Ex-ante and Ex-post.

3.3 Method and Materials

The data and Supplementary Material are available on the Open Science Framework[1].

3.3.1 Task

As our main task, I adopt the investment task by Gneezy & Potters (Gneezy & Potters, 1997). In the task, an Agent can decide the share s of her endowment to allocate to a safe asset whose return is 0 or to a risky asset whose return is +250% with probability *1/3* and -100% with probability *2/3*. The payoff of the investment task is thus

$$sE + sE \times 2.5 \times \frac{1}{3} + sE \times -1 \times \frac{2}{3} + (1 - s)E \times 1 \tag{3.1}$$

For example, if s=.5, the expected outcome of the investment is 108.3.

Given that the risky asset has an expected value of +16.7%, the investment share s is a direct measure of the risk propensity of the investor, with extremely risk-averse subjects investing s=0 and risk-neutral/seeker

[1] https://osf.io/56khw/

investing s=1.

To improve the comparability across experimental conditions, I restrict the share s to the following levels:

Table 3.1: Investment levels

0	0.25	0.50	0.75	1.0

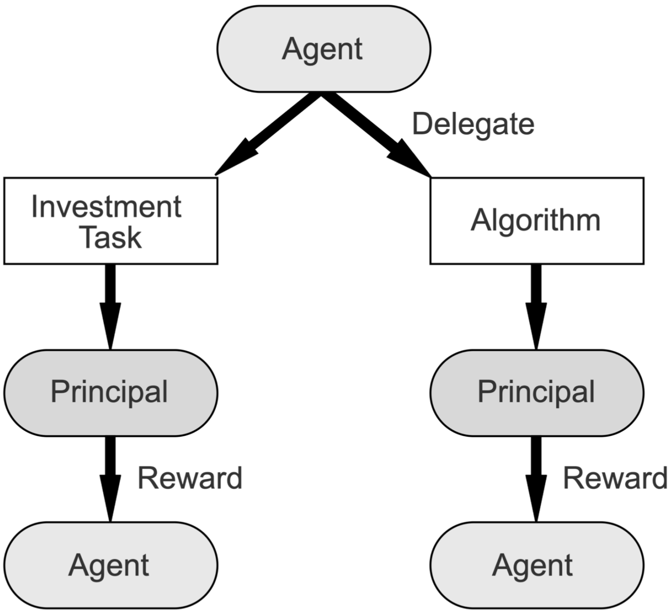
For convenience, further, in the analysis, we will use the Investment levels from Table 3.1 multiplied by 100.

Our design is an extension of Experiment 1 in König-Kersting et al., (König-Kersting et al., 2021), with two parties in the interaction and the possibility for the Agent to delegate her choice to an Algorithm (further, I will use Algo meaning Algorithm).

- Agent
  - receives E from the Principal and has to manage it
  - decides whether to invest E herself in the *Investment Task* or delegate the choice to an *Algo*
  - Gets a fixed payment, irrespective of the investment return
- Principal
  - Payoff defined by the returns of the investment task
  - Can freely punish/reward the Agent for her choice
    - ∗ The punishment/reward expressed in points from *-50* to *+50*

To define the behaviour of the Algo, I will run beforehand a **Preliminary session**: the Agent will decide sE, and the investment returns are appropriated by the Principal. The Agent will obtain a fixed payment. These sessions, which lack the reward/punishment phase, will serve as a

Figure 3.1: General experiment design



control and provide the database of choices that I will use in our Treatments. Specifically, the Algorithm in the main sessions will randomly draw investment from the distribution of choices collected in the Preliminary session.

3.4 Treatments

3.4.1 Preliminary session

In the Preliminary session, there is no Algo and no Reward/Punishment. The Principal must delegate to an Agent who makes a choice (perform Investment task).

The preliminary session runs before the main sessions (Exp 1 and Exp 2) because data from that session were collected and fed as Human Agents or Algos choices in Exp1 and Exp2.

The Preliminary session also offers a reference to check whether the behaviour of Agents and Principals differ when there is only a Human Agent.

3.4.2 Information (between-subjects)

In the main experiment (Exp 1, Exp2), I manipulate (between-subjects) the information about outcomes available to the Principal when choosing to reward or punish.

I implement the following combinations of manipulations:

Table 3.2: Dimensions that are subject to manipulation in the experiment.

Exp	1. Investment	2. Outcome	3. Delegation
1	Yes	No	Yes
2	Yes	Yes	Yes

3.4.3 Reward Before (Ex-ante) [Exp 1]

In Exp1, I have two player roles: Principal and Agent. The Agent is a Human Agent but can delegate her task to the Algo. As an Algos choice, I use randomly drawn Agents' choices that are already collected in the Preliminary session. Here I also add the Reward/Punishment conditions for participants. Specifically, the Principal has a choice to Reward or Punish the Agent before the investment results are presented (Fig.3.1). We include Punishment as a negative outcome for an Agent as an example of lossmaking (loss aversion/prospect theory).

3.4.4 Reward After (Ex-post) [Exp 2]

Exp2 is similar to Exp1, but here Principal have a chance to Reward or Punish the Agent after the results of the Investment task are presented (Fig.3.1). The order of the condition is random, and the matching among participants is random stranger across rounds.

3.4.5 Procedures and Participants

The experiment was conducted online on the Prolific[2] platform and programmed in oTree[3] (Chen, Schonger, & Wickens, 2016).

- Before beginning the experiment, participants have to read the Instructions; if they wish to proceed - accept the conditions, and conduct an attentiveness test.
- All participants are randomly assigned a role in the Investment task and perform according to the assigned role.
- Finally, they answer a survey regarding their satisfaction and concerns about investment decisions and collect demographic information.

All the interactions were managed by matching participants' choices in different roles in a "cold" way. Specifically, I collected the choices of "upstream" participants and randomly matched them with the choices of "downstream" participants. Thus, an algorithm was re-matching dropped observations. Participants did not interact directly to avoid problems typically encountered in unsupervised online experiments (no show-up, drop-off, late execution, etc.)

Participants received detailed explanations about the procedures, and all the pieces of information they receive are truthful.

Participants had to be fluent in English and be connected to a Desktop computer. Participants are also asked to answer a few attention and comprehension questions. I did not discriminate against participants on any other grounds.

In total, I recruited 901 participants (UK sample) for the experiment (Tab. 3.2). We applied "Cold" matching between choices of Agents and

[2] `https://www.prolific.co/`
[3] `https://www.otree.org/`

Table 3.3: Experiment participants structure

Condition	Principal	Agent
Preliminary session	50	50
Before - Ex-ante (Exp1)	200	201
After - Ex-post (Exp2)	200	200

Principals over distinct sessions. In the case of withdrawn participants, an algorithm was re-matching dropped observations.

3.4.6 Players

- 2 Human participants
  - **P**rincipal
    - Receives and initial endowment that is transferred to an Agent (paid fixed amount + reaward/punishment from Principal)
  - **A**gent
    - Decides on behalf of the Principal and on the money of the Principal (paid fixed amount)
- 1 **Al**go
  - Decides on behalf of the Principal on the money of the Principal

3.4.7 Payoff

Each participant receives a fixed payment from the Prolific platform for participation in the experiment. Additionally, the Principal can earn additional points from the Investment task outcome. The Agent receives a Flexible payment of up to 50 points that could be adjusted by the Principal, who rewards or punish the Agent for the choice she made.

3.4.8 Survey

After completing the main part of the experiment, participants were asked to complete a survey about satisfaction regarding the outcome of the task results, punishment or reward actions, and trust in the Agent depending on her origin. We had a different sample of questions for Agents and Principals and for the Basic and Main experiments. Survey questions can be found in *Appendix B*.

3.5 Research Hypotheses

Proceeding from the above-described, below I defined more precisely the main hypotheses:

H3.5: [Outcome bias]: *For a given investment level, higher (lower) rewards are given to Agents in **After** relative to **Before** when a positive (negative) outcome is registered.*

This hypothesis follows from the assumption that the Principal attaches a positive value to the outcomes, given a choice made by the Agent. Thus, when the outcome is positive, a positive extra reward is given to the Principal. In contrast, when the outcome is negative, an extra negative reward is given to the Principal relative to the condition in which only intentions are known.

To test this hypothesis, I will compare the rewards of Principals in *Before* (Ex-ante) and *After* (Ex-post) for a given investment level and different outcomes, controlling for delegation choices.

H3.6: [Source]: *The outcome bias depends on the source of the investment choice.*

The assumption is that outcome bias positively interacts with Agent's responsibility in the choice. Specifically, when a positive outcome is observed, a higher reward is expected for an Agent who retains the choice than for an Agent who delegates to the algorithm. In contrast, when a negative outcome is observed, a stronger punishment is expected for an Agent who retains the choice than for an Agent who delegates to the algorithm. At the same time, I expect the difference in rewards conditional upon Agent's choice to be smaller in Ex-ante than in Ex-post. This will make the outcome bias larger for choices not delegated to the algorithm.

To test this hypothesis, I will compare rewards in Ex-ante and Ex-post for a given investment level, for different outcomes, and different delegation choices of Agents.

H3.7: [Delegation]: *Agents delegate more to the algorithm in the After condition than in the Before condition.*

The Agents will anticipate that Principals will evaluate negative outcomes negatively in After (Ex-post), irrespective of investment choice, and to avoid punishment, they will alleviate their responsibility by delegating the algorithm. On the other hand, in the Before (Ex-ante) condition, Agents will retain the choice to seek rewards.

To test this hypothesis, I will compare delegation choices by Agent in Ex-ante and Ex-post.

H3.8: [Risk taking]: *Agents who retain control will make riskier investments in Before condition than in After condition.*

The Agents anticipate that Principals evaluate the investments with higher expected returns positively but evaluate bad outcomes negatively. Thus, they will seek rewards when the outcomes are not known (Ex-ante) by taking higher risks without the threat of being punished for the negative outcome.

To test this hypothesis, I will compare the investment choices of Agents who retain control over the investment in Ex-ante and Ex-post.

3.6 Analysis

3.6.1 Result 1.

Outcome bias is present as both Principal's decisions to reward or punish the Agent depend on the result of the investment task. In comparison to Ex-ante session, outcomes with winning results are rewarded, and those with unfavourable results are punished in Ex-post session.

Support: Initially, I conducted an evaluation to determine whether the investment amounts made by the Agent differed between the Ex-ante and Ex-post periods. The results of the analysis, presented in Tables 3.4 and 3.5 indicate that there was no statistically significant difference observed between Ex-ante and Ex-post investment amounts (Wilcoxon rank sum test with continuity correction, p-value > 0.05).

Table 3.4: Ex-ante and Ex-post investment

player.treatment	**Mean**	**Median**	**SD**
Ex-post	43.25000	50	25.02637
Ex-ante	42.78607	50	26.99906

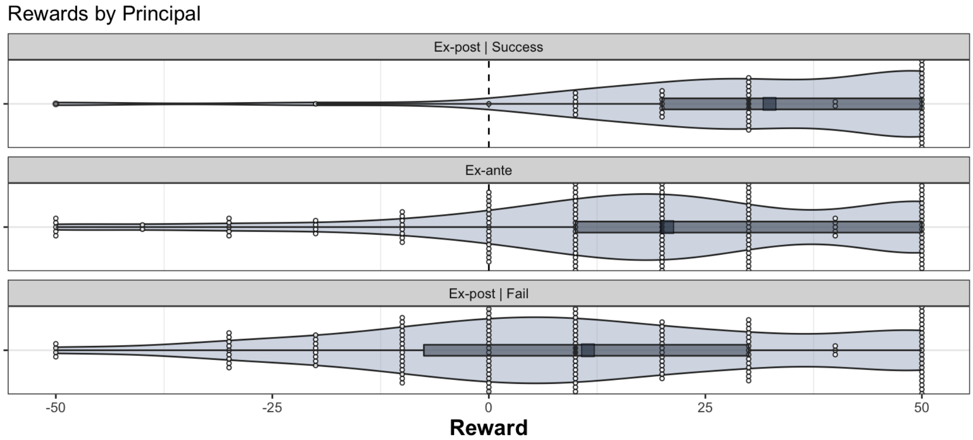
Fig 3.2 compares the following distributions of Principal choices to reward or punish agents:

- when Principal knew the investment outcome and this outcome was

Table 3.5: Ex-ante and Ex-post investment - Wilcoxon test

statistic	p.value	method	alternative
19791	0.779728	Wilcoxon rank sum test with continuity correction	two.sided

Figure 3.2: Rewards in Ex-post|Success, Ex-ante, and Ex-post|Fail



successful (Ex-post|Success)

- when Principal did not know the investment outcome (Ex-ante)
- when Principal knew the investment outcome and the outcome was a failure (Ex-post|Fail)

A statistically significant differences was observed between the Ex-ante and Ex-post periods for both the Ex-post|Fail and Ex-post|Success outcomes (Wilcoxon rank sum test with continuity correction; p-value = 0.0002 for Ex-post|Fail and p-value = 0.0004 for Ex-post|Success). Additionally, a statistically significant difference was observed between Ex-post|Fail and Ex-post|Success (Wilcoxon rank sum test with continuity correction, p-value $\ll 0.05$).

To further investigate, a linear regression model was employed, with Agent reward as the dependent variable and Ex-post|Fail and Ex-post|Success as independent variables:

$$Agent\ reward \sim \alpha + \beta_1 Ex\text{-}post|Fail + \beta_2 Ex\text{-}post|Success + u \qquad (3.2)$$

Table 3.6: Reward by Ex-post|Success, Ex-ante, Ex-post|fail regression

	(1)
(Intercept)	20.597 (1.728)***
Ex-post|Success	11.810 (3.755)**
Ex-post|Fail	-9.159 (2.664)***
R^2	0.072
Adj. R^2	0.067
Num. obs.	401

$p < 0.001$; $p < 0.01$; $p < 0.05$; $\cdot p < 0.1$

Table 3.6 presents the regression parameters. Upon examination, a positive statistically significant impact of Ex-post|Success (11.810, p-value<0.01) was observed, while a negative statistically significant impact of Ex-post|Fail (-9.159, p-value<0.001) was noted. Moreover, the Ex-post|Success coefficient exhibited a greater absolute value than the Ex-post|Fail coefficient, indicating that Principals tended to give higher rewards to the Agents for successful investments and punish less for investment failures. Furthermore, the positive intercept (p-value<0.001) suggests that by default, the experiment participants were predisposed towards giving rewards to Agents.

Figure 3.3 provides a comparison of rewards Ex-ante, Ex-post|success, Ex-post|fail for each investment level (Table 3.1). Notably, the earlier observed outcome bias persist in the case of Ex-post|success, with extreme investment levels such as 100, 75, and 0 being associated with a higher amount of rewards received by Agents as compared to Ex-ante. Upon analysing Ex-post|Fail, a similar outcome bias was observed for investment levels of 100 and 75, with the amount of rewards received by Agents being lower in Ex-post|Fail when compared to Ex-ante.

Figure 3.3: Rewards in Ex-post|Success, Ex-ante, and Ex-post|Fail depending on investment level

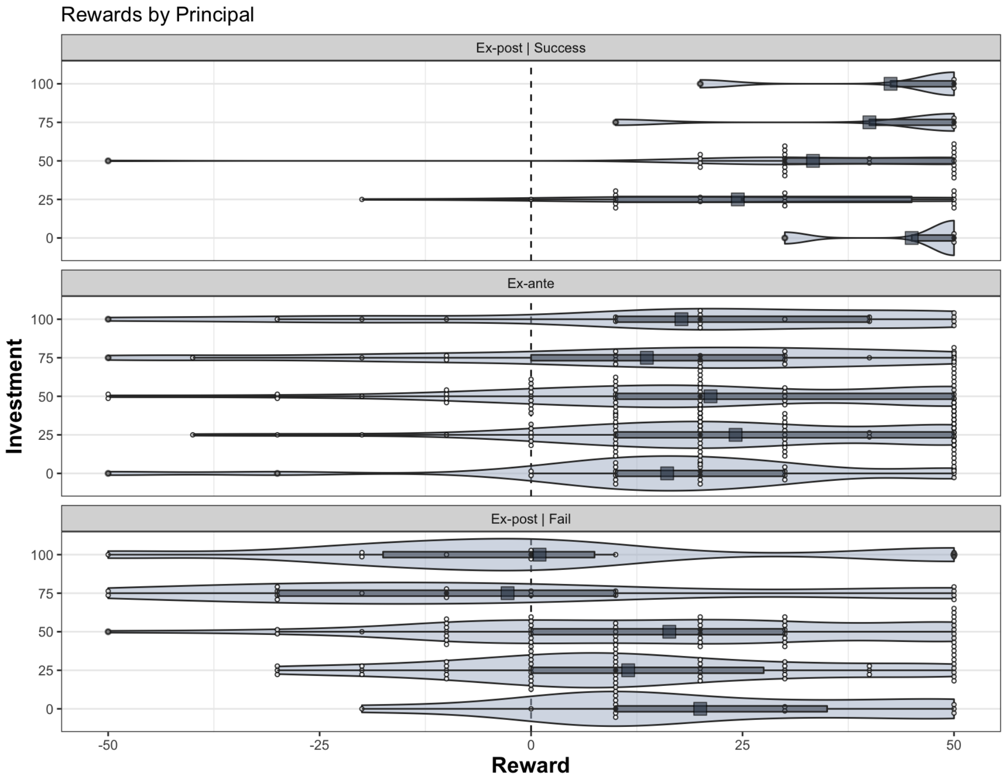


Table 3.7: Principal reward in Ex-post Success and Ex-post Fail regression

	Model 1
player.invest	**0.332 (0.036)*****
fail_after	**19.226 (4.231)*****
success_after	**27.096 (7.244)*****
player.invest:fail_after	**-0.512 (0.092)*****
player.invest:success_after	-0.210 (0.150)
R2	0.340
Adj. R2	0.331
Num. obs.	401
p <0.001; p <0.01; p <0.05; ·p <0.1	

Conclusion: In light of the analysis comparing the Principals' decision to Reward or Punish Agent in Ex-ante and Ex-post for given investment levels and different outcomes, it can be concluded that H3.5 is supported.

3.6.2 Result 2.

The outcome bias does not depend on the source of the investment choice. There are no statistically significant differences in the Principal's decision to reward or punish the Agent despite her choice to delegate to the Algo or invest on her own.

Support: We compared rewards in delegated and not delegated cases in Ex-ante and Ex-post (Figure 3.4). Results show that there is no statistically significant difference between the Agent's choice to invest herself or delegate investment to Algo:

- Ex-post|Fail: p-value = 0.220
- Ex-post|Success: p-value = 0.827
- Ex-ante: p-value = 0.245

Figure 3.4: Reward of Algo and human in Ex-post and Ex-ante

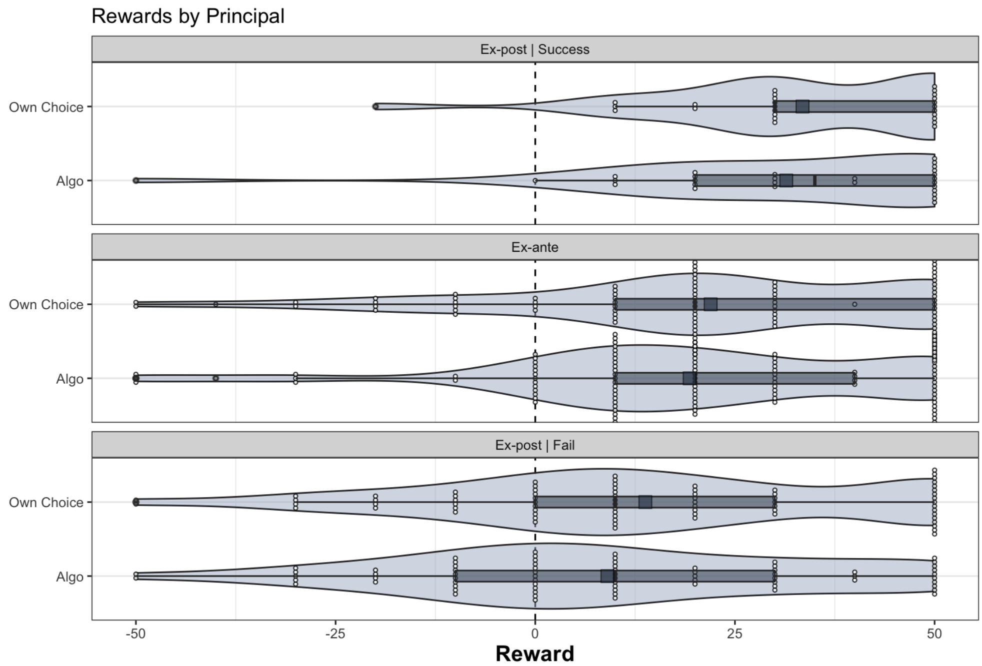


Figure 3.5 shows rewards by investment level. When an Agent decides to make an investment on her own in Ex-post|Fail, lower rewards correspond to higher investment levels. For the highest investment levels (75 and 100), I even observe negative rewards (e.g., punishment). At the same time, when Agents delegated to the Algo and investment failed, Principals still rewarded the Agents. The reward amount did not depend on the investment level and was positive, except when the investment level equalled 75).

For the Ex-post|Success, when an Agent delegated investment choice to the Algo, I observed a slightly positive trend: the higher the investment level of the Algo, the bigger the reward (with an exception when the investment was equal to 0). On the other hand, when the Agent invested herself, I did not observe a similar trend in the amount of reward.

Table 3.8: Principal reward in Ex-post|Success and Ex-post|Fail regression

	(1)	**(2)**
(Intercept)	**19.327 (2.405)*** **	**16.087 (5.059)** **
Success|Ex-post	**12.102 (5.223)***	**28.913 (13.143)***
player.delegateOwn Choice	2.632 (3.463)	
Fail|Ex-post	**-10.299 (3.761)** **	3.913 (8.640)
Success|Ex-post:player.delegateOwn Choice	-0.599 (7.525)	
player.delegateOwn Choice:Fail|Ex-post	2.124 (5.337)	
as_factor(player.invest)25		8.092 (5.863)
as_factor(player.invest)50		5.129 (5.792)
as_factor(player.invest)75		-2.403 (7.522)
as_factor(player.invest)100		1.691 (7.635)
Success|Ex-post:as_factor(player.invest)25		-28.648 (14.637)·
Success|Ex-post:as_factor(player.invest)50		-16.796 (14.326)
Success|Ex-post:as_factor(player.invest)75		-2.597 (18.732)
Success|Ex-post:as_factor(player.invest)100		-4.191 (18.778)
as_factor(player.invest)25:Fail|Ex-post		-16.611 (9.712)·
as_factor(player.invest)50:Fail|Ex-post		-8.783 (9.691)
as_factor(player.invest)75:Fail|Ex-post		-20.375 (11.761)·
as_factor(player.invest)100:Fail|Ex-post		-20.691 (12.892)
R^2	0.077	0.118
Adj. R^2	0.065	0.086
Num. obs.	401	401

p <0.001; p <0.01; p <0.05; ·p <0.1

I also ran regression confirms that delegation to Algo has no significant impact on Agent Reward compared to the option to perform the investment personally.

Figure 3.5: Principal reward in Ex-post Success and Ex-post Fail by investment level

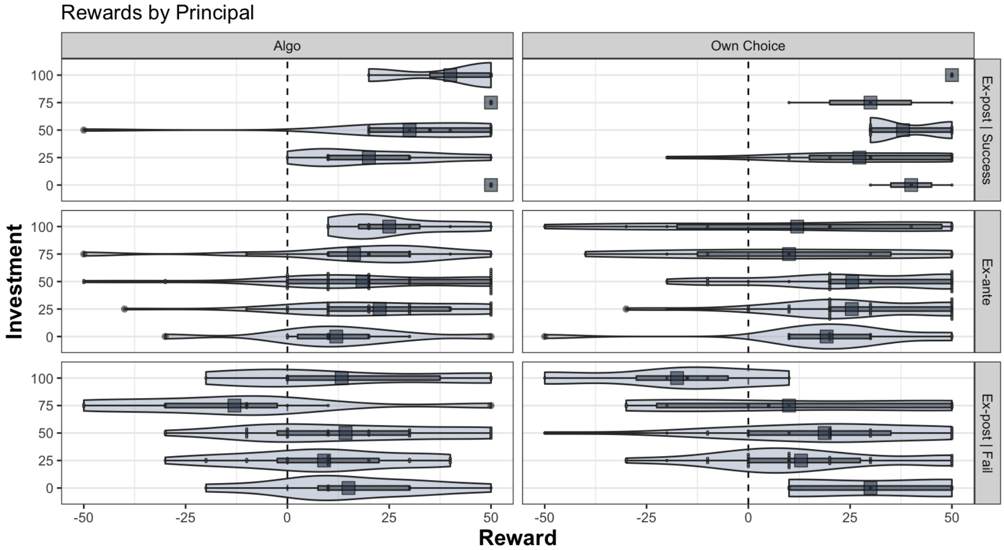


Table 3.9: Principal Reward in Ex-post|Success and Ex-post|Fail regression by investment level

	0	**25**	**50**	**75**	**100**
(Intercept)	12.000 (7.018)·	**22.667(3.904)*****	**18.444(3.681)*****	16.364 (9.211)·	**25.000(9.944)***
Success|Ex-post	**38.000(17.190)***	-2.667 (8.975)	11.556 (7.557)	33.636 (23.484)	15.000 (19.042)
player.delegateOwn Choice	7.231 (9.335)	2.739 (5.253)	7.073 (5.881)	-6.364 (14.195)	-13.000 (13.341)
Fail|Ex-post	3.000 (10.527)	**-13.667(6.173)***	-4.159 (5.944)	**-29.364(13.348)***	-11.667 (15.190)
Success|Ex-post:player.delegateOwn Choice	-17.231 (24.075)	4.534 (11.597)	0.927 (11.795)	-13.636 (33.686)	23.000 (35.111)
player.delegateOwn Choice:Fail|Ex-post	7.769 (16.487)	1.202 (7.994)	-2.609 (9.043)	29.364 (20.285)	-17.833 (22.530)
R^2	0.191	0.089	0.067	0.229	0.286
Adj. R^2	0.069	0.054	0.034	0.119	0.148
Num. obs.	39	139	150	41	32

p <0.001; p <0.01; p <0.05; ·p <0.1

As Fig.3.5 shows, even if the source of investment choice does not affect outcome bias, I can still see that when the Principal knew the result of the investment task and the Agent invested on her own, the Principal punished the Agent more with the increase of the invested (lost) amount. I do not observe this pattern for the same treatment when the Agent delegated investment choice to the Algo. In fact, I can say that there was practically no difference in reward with regard to investment amount (except for the 75 investment level).

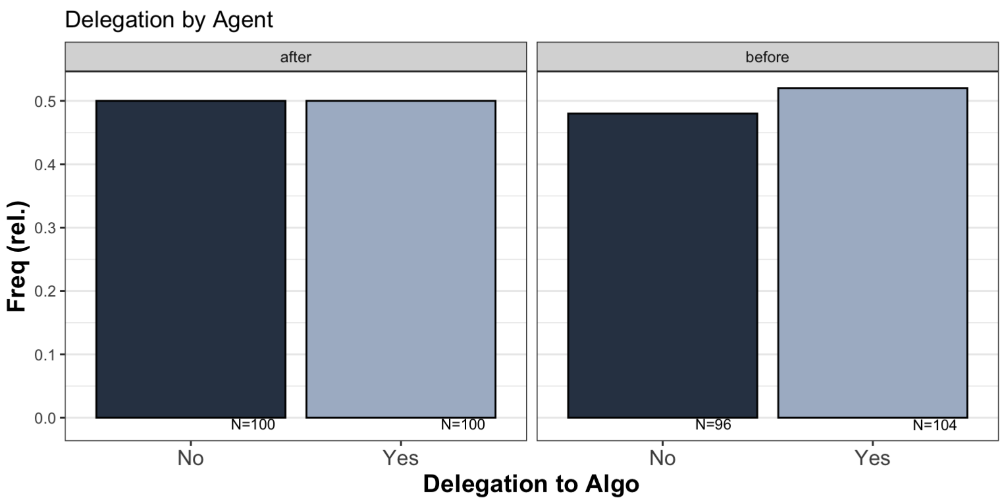
Conclusion: After comparing the amounts of Rewards/Punishments that Principals granted to the Agents based on their choice to invest themselves or to delegate to the Algo, I can not confirm the H3.6.

3.6.3 Result 3.

There is no significant difference in delegation between Ex-ante and Ex-post conditions. Agents do not delegate more to the Algo in the Ex-post condition than the Ex-ante condition.

Support: Fig 3.6 shows the delegation choices of an Agent. In the Ex-post condition, I observed that half of the participants delegated to the Algo and half invested on their own. In Ex-ante, 48 did not delegate, while 52 delegated to the Algo. Results show that the difference in the delegation choices of an Agent between Ex-ante and Ex-post conditions is not statistically significant (Chi-squared test, p value>0.05). This indicates that Agents were not affected by the potential unfavourable evaluative reaction from the Principal.

Figure 3.6: Delegation choices in Ex-ante and Ex-post



However, comparing age brackets older than 40 and younger than 40, in the Ex-ante condition, older participants delegated much more than younger agents. In other words, in Ex-ante younger participants preferred not to delegate, while older participants chose more to delegate than not to delegate.

Conclusion: After comparing delegation choices of Agents in Ex-ante and Ex-post conditions, I can not confirm the H3.7.

3.6.4 Result 4.

The investment does not differ across conditions for the Agents who retain control. Agents who retain control do not make a riskier investment in the Ex-ante condition than in the Ex-post condition.

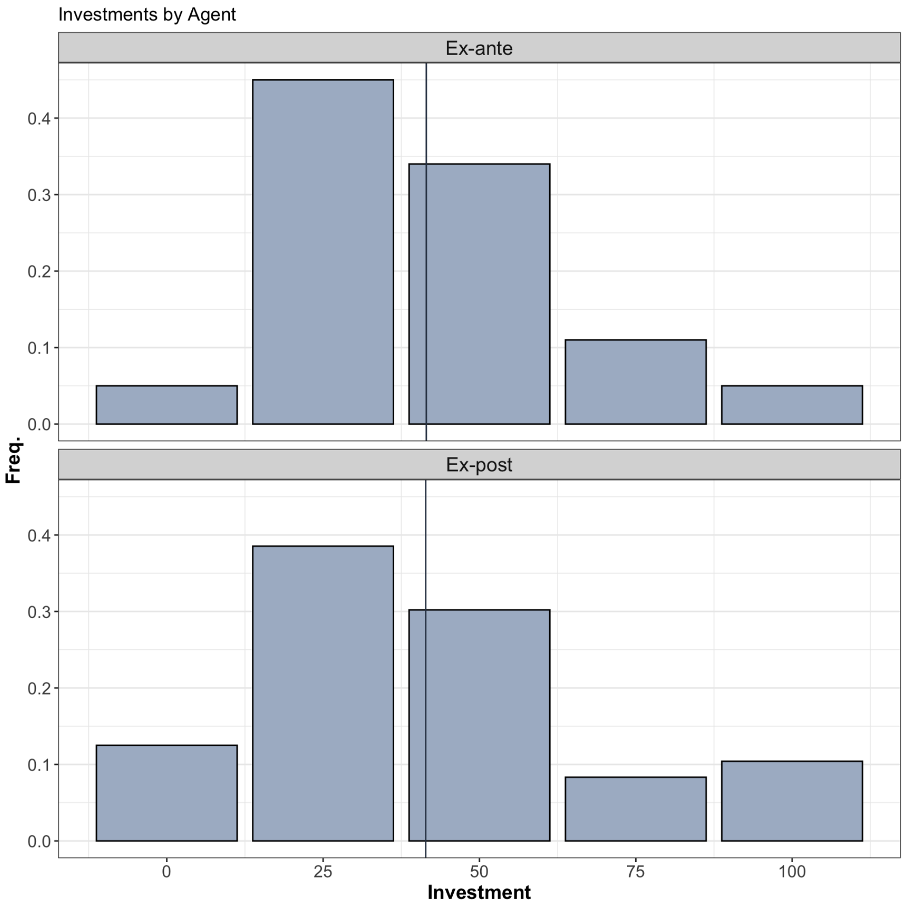
Support: We compared the investment choices of Agents who retain control over the investment in Ex-ante and Ex-post (Fig 3.7). The investments do not differ across conditions for the Agents who invested on their own (retained control), so I found no difference in the investment level among treatments. Thus the findings can not support H3.8.

Survey. After the main part of the experiment, participants conducted a short survey both for the Preliminary session and the main part.

Agents. On the statement that Agents would receive more tokens from the Principal if they would delegate to the Algo, those Agents who retain control mostly disagreed; in contrast, those who did delegate to Algo had a positive opinion both for success and fail conditions. This confirms their delegation choice and the belief that this will somehow positively reflect on their reward.

Those agents who delegated to the Algo agreed and strongly agreed that their primary motive was to avoid losses. While the Agents who made the decision themself although agreed with the statement, however in the event

Figure 3.7: Investment level in Ex-post and Ex-ante



of a loss, a sufficiently large percentage of the agents did not agree that they wanted to avoid losses. I suggest that this demonstrates a desire to justify the actions: "I delegate to the Algorithm - I only wanted to avoid losses"; "if I did not delegate and lost the investment, I did not seek to avoid losses".

Principals. When the investment failed, Principals agreed with the statement that they would give more Tokens to the Agent if she/he made a different delegation choice (both when the Agent delegated to the Algo and did not).

More robust results for the statement "How much I will reward/punish the Agent depends on whether I know the result of the investment task or not" in no delegation choice. Principals also believed that Agent would make a different investment if there was no reward/punishment, and he/she should feel accountable for the investment outcome.

For the statement: "I believe that, in general, an algorithm would make better investment decisions than a human" in success case, positive and negative answers are in balance, while when investment failed, there are more positive answers.

Questions and additional materials can be found in *Appendix B*.

3.7 Discussion

Outcome bias. In this chapter, it was observed that most rewards in Ex-ante were positive and did not depend on the investment amount. This might be due to a generally positive attitude of Principals: their decision to reward or punish was made before they knew the results of the investment task. It is also possible to assume that the Principals believed that regardless of the outcome, the agents deserved a reward for the task accomplished. In Ex-post|Success, I observe that the reward also increases with an increase in investment amount. However, when the Agent chose not to invest in a risky asset (investment = 0), I observed that the Agent received the highest reward. In Ex-post|Fail, I observe a mirrored trend compared to Ex-post|Success: the higher the investment, the lower the reward from the Principal. For investment levels 75 and 100, I observe many punishment decisions.

Thus, I observe a clear relationship between the investment outcome and rewarding the Agent for her decision. When the outcome was positive, a positive extra reward was given to the Agent by the Principal. In contrast, when the outcome was negative, an extra negative reward was given to the Agent relative to the condition in which only intentions were known. The high rewards for keeping the safe option (no investment) could be explained by the fact that in the case of non-investment and investment success, the Principals were relatively calm to this. In the case of non-investment and failure, the Principals were satisfied that the Agent did not invest and thus kept the original endowment.

The outcome bias does not differ when a choice is made personally or delegated to the algorithm. Additionally, delegation to an algorithm and the level of risky investments do not differ when knowing that rewards or punishments are given before or after the investment outcome. This shows that the Principal's decision to reward or punish is driven more by the outcome rather than the source (Human Agent vs Algo); this might be due to the Principal's interest in receiving a higher outcome rather than caring much about who is performing the investment task.

According to Result 3.6.2, the source of investment choice does not influence outcome bias. In other words, as the Agent is Rewarded/Punished conditional upon outcomes also, when the machine chooses, the outcome bias is not mediated by delegation to an Algo. However, a human agent was punished more with an increased invested (lost) amount. This suggests that an Agent was better treated when she delegated her choice to an Algo, and the investment failed. This suggests that humans and algorithms are treated differently. Since the punishment was severer for the higher investments, in order to avoid punishment, the riskier the stake - the better to use algorithms as an executor.

Anticipates the Principal's evaluation. It is not possible to support the idea that an Agent anticipates the Principal's evaluation regarding investment outcome. Thus, Agents did not seek rewards when the outcomes were unknown (Ex-ante) by taking higher risks without the threat of being punished for the negative outcome.

It might be assumed that Agent was not concerned by the information available to the Principal, or she did not anticipate the perceived shift in responsibility within the delegation and was not focused on minimizing the downward of it. Either she was not concerned about potential reward/punishment or did not believe she might avoid it. There is also an option: the Agent's punishment and reward were so small that they had no weight in the Agent's decision-making, so the Agent ignored them.

The current finding contradicts with results of Holzmeister et al. (Holzmeister et al., 2022), who found a correlation between delegation rates and the possibility of delegating to the Algo. This might be due to a different design of the experiment, and in our paper, the main focus was the possibility of delegation to an algorithm, not another human, with the following impact of the delegation on agents' reward. Specifically, in our experiment, Agents had a choice between delegation to the Algo or performing the task herself, while in the paper of Holzmeister (Holzmeister et al., 2022), participants had a chance to delegate to the other human being(finance professional). In contrast, our results demonstrate that the outcome bias does not differ when a choice is made personally or delegated.

Delegation to an Algo does not differ when knowing that rewards are given before or after the investment outcome. In other words, the Agents do not tend to "hide" behind the machine. There is also no shift in responsibility to save on self-image concerns. We assume that the reason is similar to the Principal's - the Agent is concerned about the reward/punishment from the Principal. Further research (possibly qualitative) is needed regarding the motivations of Agents.

These findings augment understanding of the delegation preferences regarding interaction partner choice.

3.8 Conclusion

In this chapter, I examined how the information about the investment's outcome affects the human incentive behaviour depending on the interacting partner. In the experiment, the Principal delegated the Agent a risky financial decision to invest all or a part of the Principal's endowment. The Agent had a choice to solve the investment task on her own or to delegate the investment task to the Algorithm (Algo). The Principal, in turn, rewarded or punished the Agent for her choice. (Sentence to describe Ex-ante and Ex-post)

In line with the literature, my findings largely confirm the presence of outcome bias and the influence of a random outcome of the risky investment on the Principal's evaluations and rewards for the Agent. I demonstrate that relative to the Ex-ante, winning outcomes get rewarded and losing outcomes get punished. The outcome bias does not depend on the source of the investment choice. Agents do not delegate more to the algorithm in the Ex-post condition than in the Ex-ante condition. Furthermore, those Agents who retain control do not make riskier investments in the Ex-ante condition than in the Ex-post condition.

The current experiment represents a practical case and findings have implications for growing financial markets as human dependence on algorithms increases in various fields, including finance. Our study highlights the potential pathways that need to be considered when designing human-algorithm interactions. Firstly, we demonstrate the presence of outcome bias in investment decisions made by artificial agents. Thus, future research should consent rate on minimisation of bias in investment decision-making processes as much as possible. Secondly, our study reveals that individuals in high-risk environments do not exhibit a distinct preference for using algorithms over other humans to deflect responsibility or shift the blame. This suggests that while developing a framework for human-machine interactions, individuals can entrust both human and algorithm agents with responsibility and avoid using algorithms as scapegoats. Finally, despite the rise of robo-advisers, recipients of the decision-making and responsibility attribution will still be held by a human. Considering the systematic analysis of human interactions with algorithms can enhance our understanding of assigning responsibility to respective parties.

Conclusions

In this thesis, relevant human and machine interaction issues were raised and considered. The examination and replenishment of knowledge in the field of interaction between humans and robots or smart assistants is a paramount issue in the current conditions of society and technology development. My goal was to contribute to investigating the interaction between humans and technology, in their characteristics, problems and potential threats.

The first chapter urges regulators and policy-makers to implement legal guidelines for using AI in financial decision-making so that the outcome can be beneficial for those most in need. Two conditions suggested to hold to achieve the most beneficial outcome in interactions with technology. Firstly, data used for targeting and enforcing social protection programs should be exhaustive and include all ranges of social and economic layers of the population. Additionally, depending on the system's characteristics and particular circumstances, individuals must be given the possibility to consciously decide to rely on system or not. Furthermore, human cognitive offloading and its impact on behaviour should be studied thoroughly before implementing specific policy initiatives. Future research across the behavioural sciences should thus aim to comprehensively investigate how specific problems related to financial decision-making under scarcity might be alleviated by using AI, particularly how that might be done without putting the individual at increased risk. Studying further and developing practical implementations of how AI could aid financial decision-making under economic scarcity is necessary. This form of research will benefit not only the ones with the least available resources but society as a whole.

The second chapter demonstrates that in offloading cognitive tasks on algorithms, humans should be careful because blindly following AI suggestions, one can obtain a very high score in one case but fail dramatically in the other depending on AI accuracy. However, results showed that participants offload cognitive tasks on AI even if it is not accurate and safe. Therefore, AI presence negatively impacts overall human performance. This is a potential threat to human-technology interaction. Therefore, this chapter contributes to the literature by providing additional insights concerning human technology collaboration. Future research should consider the broader meanings of time and penalties and the impact of the prolonged effect on human behaviour of various social strata.

Finally, the third chapter answered how the information about the investment's outcome affects the human incentive behaviour depending on the interacting partner. In line with the literature, my findings largely confirm the presence of outcome bias and the influence of a random outcome of the risky investment on the Principal's evaluations and rewards for the Agent. The outcome bias, though, does not depend on the source of the investment choice. Agents do not delegate more to the algorithm in the Ex-post condition than in the Ex-ante condition. Furthermore, those Agents who retain control do not make riskier investments in the Ex-ante condition than in the Ex-post condition. The current experiment contributes to understanding human-machine interaction and human incentive behaviour depending on the interacting partner. Thus, the results may have implications for today's markets, as robo-advice recipients will still be responsible for promoting them.

The current thesis contributes to the study of human-machine relations, considering it from three different angles such as the impact and potential assistance to certain social strata with the help of algorithms, the effect of factors such as time and penalty on the decision-making process involving algorithms, and the delegation of financial decisions to human and algorithm.

Appendix A

A.1 Survey

The Likert scale was used from strongly disagree to Strongly agree for Attitude questions.

Survey: Your attitude towards AI and information technology.

1. I think AI could solve precision/computational tasks better than humans.

2. I think that AI is better than humans in problem-solving tasks.

3. I use information technology only to support my decisions.

4. When you are under time pressure, smart assistants hinder more than help.

5. I usually trust technology until it gives me a reason not to.

6. I tend to trust information technology when I first use it.

7. I cannot trust technology when it comes to high financial risks.

8. In general, I will continue to use smart assistance and information technologies, even if it fails me several times.

A.2 Survey responses before and after the experiment

Here we can see the distribution of answers depending on the presence and absence of AI assistance and before/after conditions.

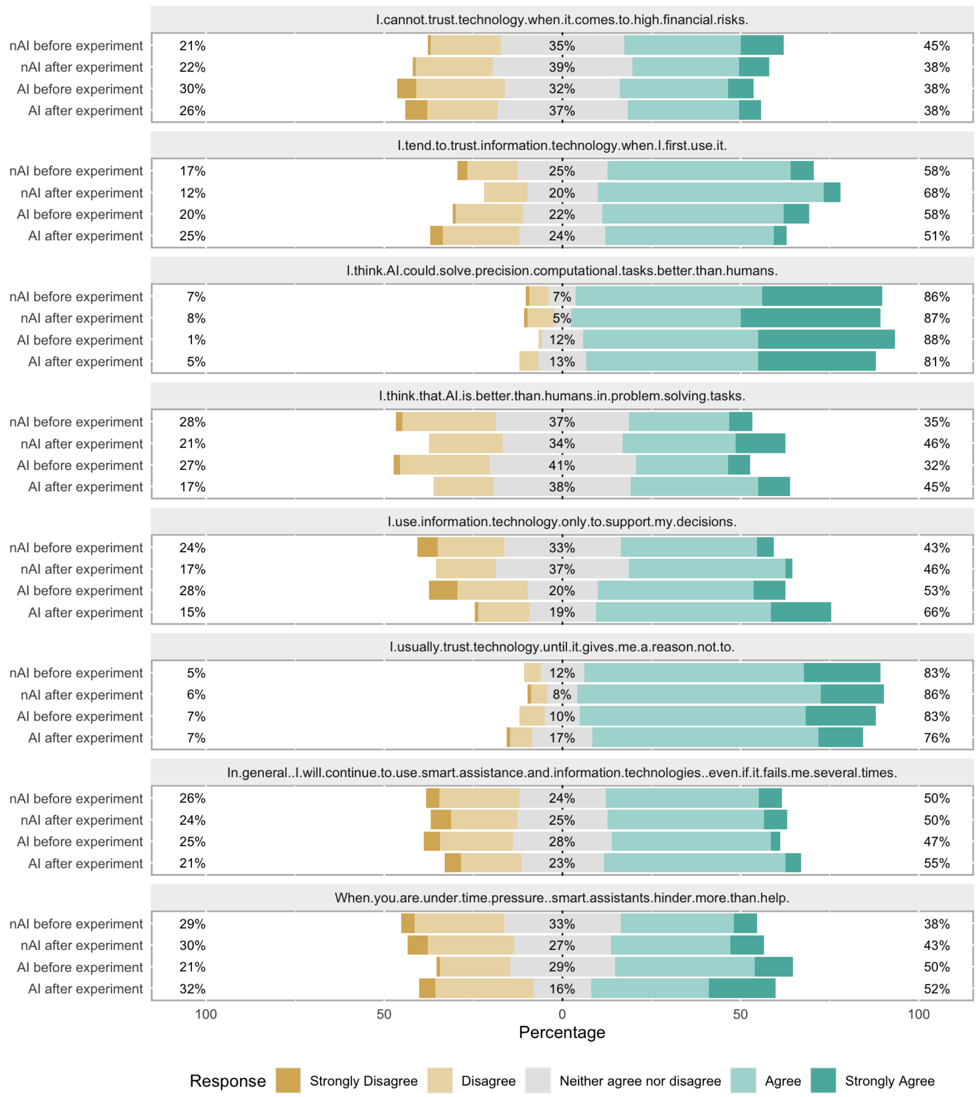


Figure A.1: Survey responses before and after the experiment

A.3 Second part.

A.3.1 Instructions for the second part

Visual comparison task.

You will have 4 rounds of choice with 10 trials of visual comparison task in each round. In each trial, you have to compare 2 objects on the screen. The objects will look like this Changing shape, size, and mirroring is considered a difference; rotation is not considered a difference. Before starting the actual task, you are given 3 trial rounds to become acquainted with the task.

AI Help. (only available in Treatments session)
In each trial, there will be a simulated AI assistant that will suggest you the answer with a flashing button like this The simulated AI assistant is not perfect but improves round by round. Therefore, the probability of the correct suggestions will be the following:

Table A.1: The probability of the correct suggestions

Round	Accuracy
1	50% (5 out of 10)
2	70% (7 out of 10)
3	80% (8 out of 10)
4	90% (9 out of 10)

Payoff and Endowment.
Your initial endowment is 400 points, which is equivalent to 4 GBP. Thus every point is worth 0.01 GBP (=1p). Each correct answer will allow you to keep your endowment safe. An incorrect answer will fine you the low penalty (-1 point) or high penalty (-10 points), you will be informed in advance if the round is low-penalty or high-penalty. Your final bonus payment in GBP is given by the initial endowment minus the total penalty.

Timing.

The whole experiment will take around 15-20 minutes. Once you start the experiment, you will be able to take breaks between rounds, but not during the round. For each trial, you have a maximum of 10 (for highTime-C, and 17 for Accuracy performance goal) seconds to answer.

Control questions.

Before continuing, please answer these questions:

- *How many seconds do you have to answer for each trial?*
- *Is rotation considered as a difference?*

A.3.2 Training

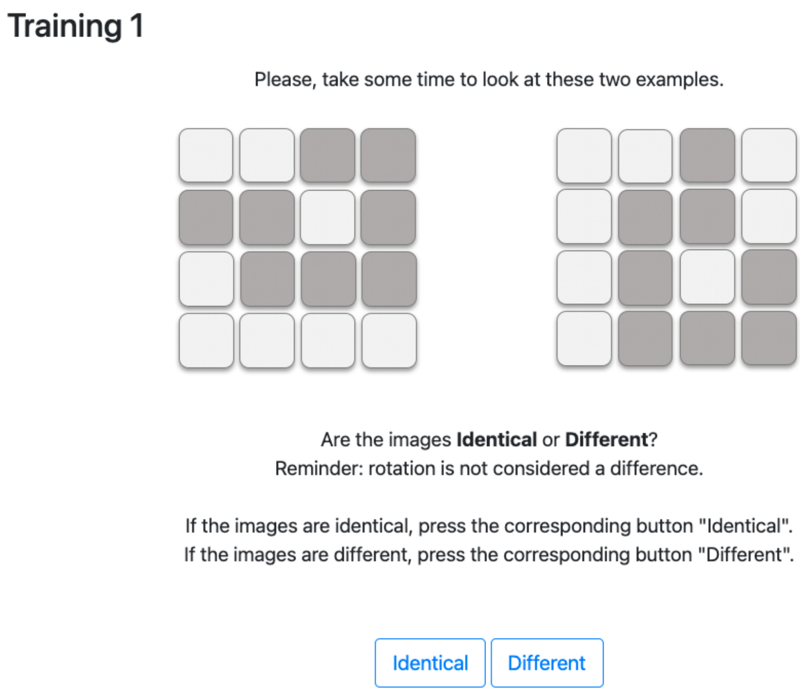


Figure A.2: Example of a training task

If participant give correct answer he received positive confirmatory comment (Fig. A.3).

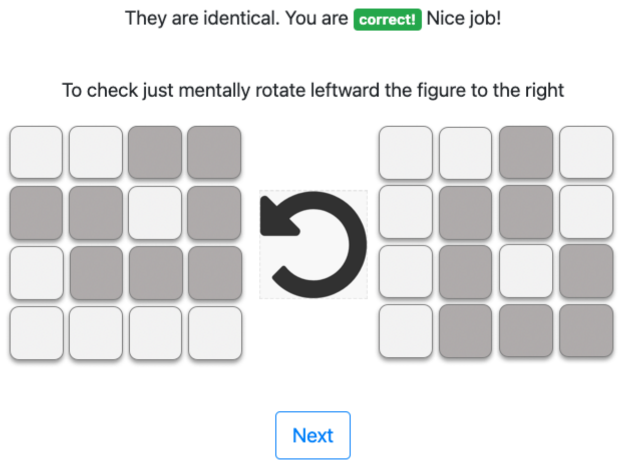


Figure A.3: Example of a feedback on training task

A.4 Third part

A.4.1 Survey 2

The Likert scale was used from strongly disagree to Strongly agree for Attitude questions.

Survey: Your attitude towards AI and information technology

1. I think AI could solve precision/computational tasks better than humans.
2. I think that AI is better than humans in problem-solving tasks.
3. I use information technology only to support my decisions.
4. When you are under time pressure, smart assistants hinder more than help.
5. I usually trust technology until it gives me a reason not to.
6. I tend to trust information technology when I first use it.

7. This question is just to check your attention. Please, click Strongly disagree.

8. I cannot trust technology when it comes to high financial risks.

9. In general, I will continue to use smart assistance and information technologies, even if it fails me several times.

Demographic questions

1. What gender do you identify as?: *Female, Male, other, Prefer not to answer.*

2. What is your age? *(drop-down list)*

3. What is the highest degree or level of education you have completed? (Some High School, High School, Bachelor's Degree, Master's Degree, Ph.D. or higher, Trade School, Prefer not to say)

4. What is your current employment status? (Employed Full-Time, Employed Part-Time, Seeking opportunities, Retired, Prefer not to say)

5. What is your annual household income? (Less than 25.000, 25.000 - 50.000, 50.000 - 100.000, 100.000 - 200.000, More than 200,000, Prefer not to say)

6. What is your country of residence?

Appendix B

B.1 Questionnaire

Preliminary session - Agent

1. I care about the outcome in the Investment Task.
2. I deliberately chose the amount invested in Asset A.
3. My main motivation for the investment choice was to avoid losses.
4. I made my investment choice randomly.
5. My main motivation for the investment was to multiply the endowment.
6. I feel no accountability to the Principal for the investment outcome.

Preliminary session - Principal

1. I am satisfied with the results.
2. I believe Agents should feel accountable for negative investment outcomes.
3. I am satisfied with Agentś choice.
4. I believe that, in general, an algorithm would make better investment decisions than a human.

5. I believe Agents should feel accountable for positive investment outcomes.

Main Agent

1. I believe I would receive more Tokens from the Principal if I delegated to the Algorithm.
2. If the Principal knows the investment results before deciding to reward/punish, it is better to delegate the task to the algorithm.
3. I could make a riskier investment if reward/punishment happens before the Principal knows the results.
4. My main motivation for the investment choice was to avoid losses.
5. I believe the Principal would reward/punish differently before or after the results are known.
6. I would make a different investment if there was no reward/punishment.

Main Principal

1. I would give more Tokens to the Agent if she/he made a different delegation choice.
2. How much I will reward / punish the Agent depends on whether I know the result of the investment task or not.
3. I believe the Agent would make a different investment if there was no reward/punishment.
4. I believe Agents should feel accountable for investment outcomes.
5. I believe that, in general, an algorithm would make better investment decisions than a human.

6. I believe Agents should feel accountable for investment outcomes, even if she/he delegate to the Algorithm.

Visualisation of the participants' answers:

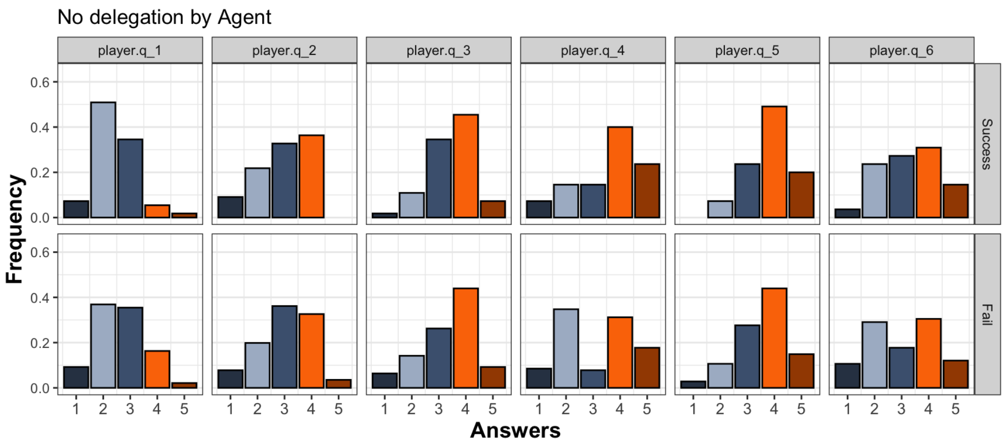


Figure B.1: No delegation to Algo

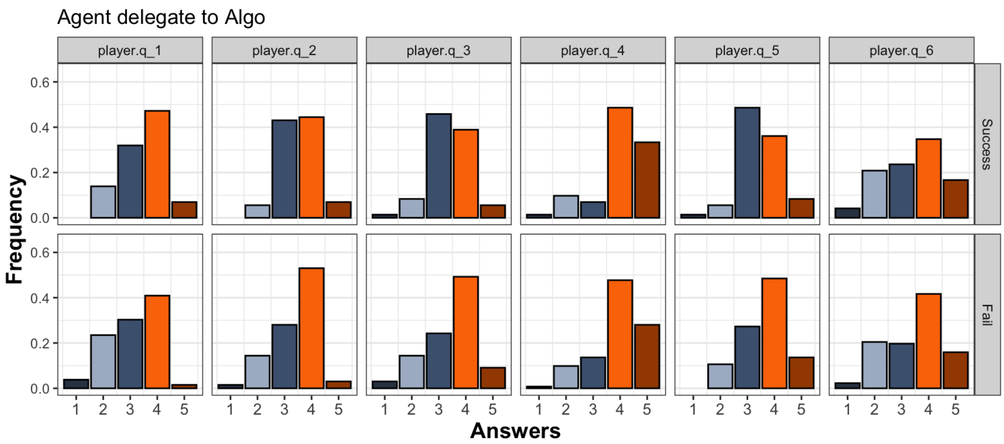


Figure B.2: Agent delegate to Algo

Comparing age brackets older than 40 and younger than 40, in the Ex-ante condition, older participants delegated much more than younger agents. In other words, in Ex-ante younger participants preferred not to delegate, while older participants chose more to delegate than not to delegate (See Fig.B.6 and Fig.B.5).

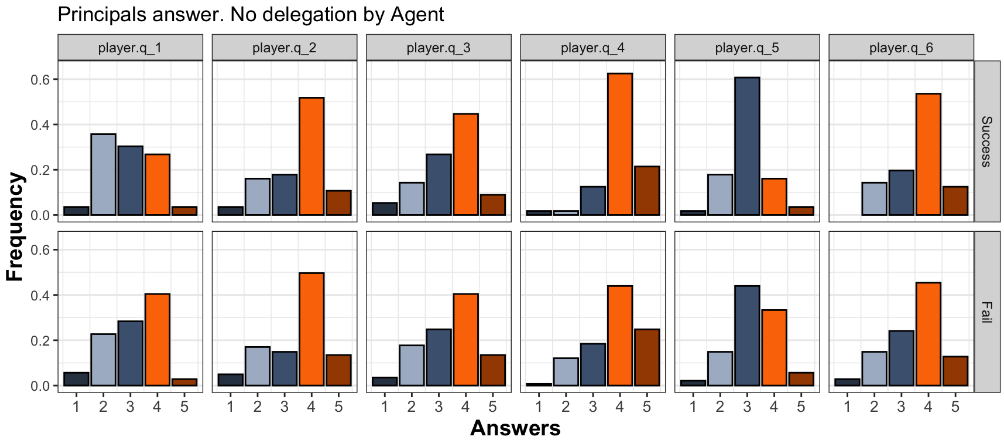


Figure B.3: Principals answer. No delegation to Algo

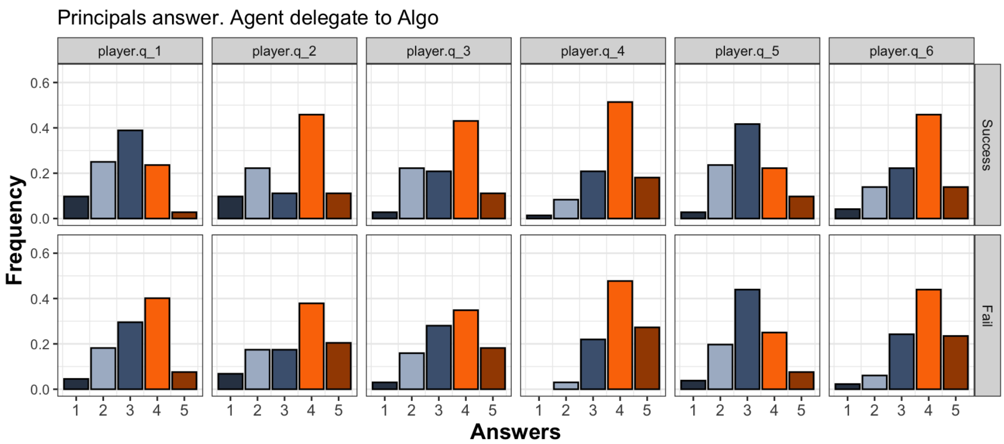


Figure B.4: Principals answer. Agent delegate to Algo

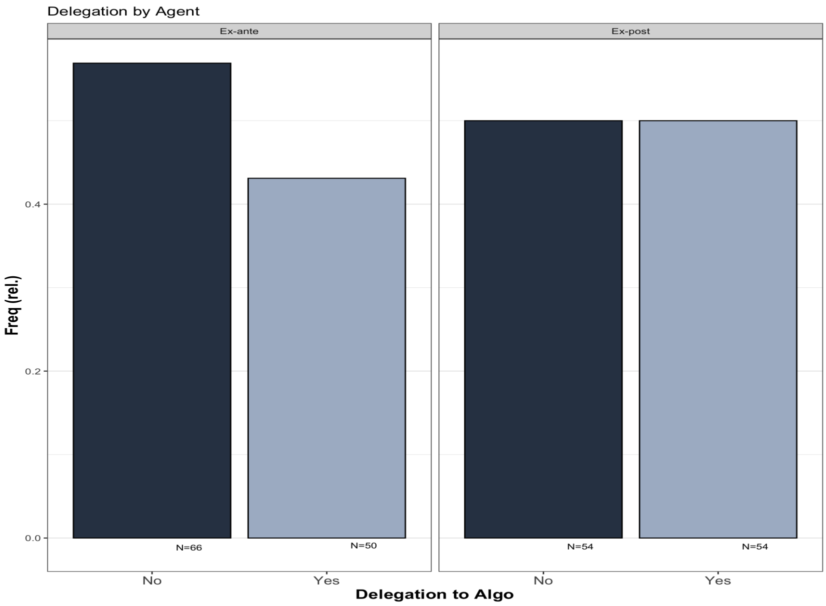


Figure B.5: Delegation by Agent, age less or equal to 40

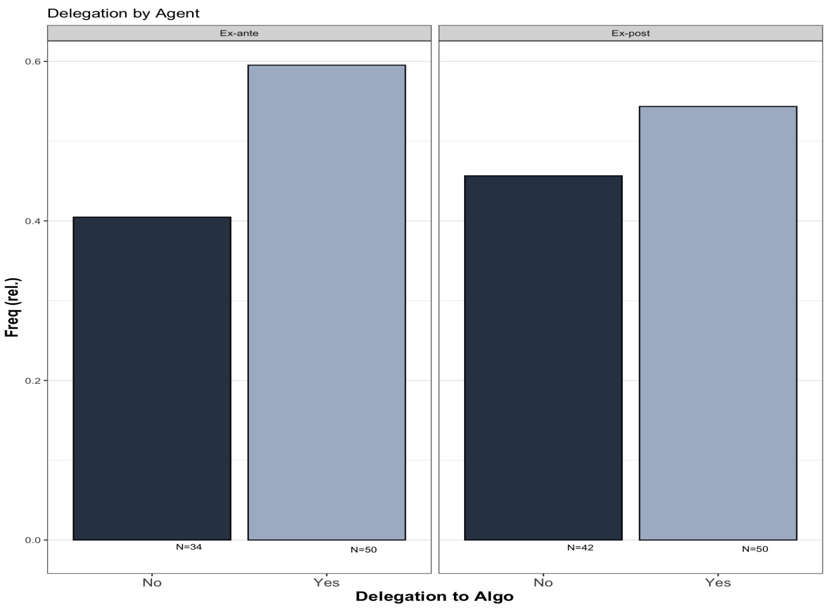


Figure B.6: Delegation by Agent, age more 40

Bibliography

Accenture. (2017). *Global distribution and marketing consumer study.* https://www.accenture.com/_acnmedia/Accenture/next-gen-3/DandM-Global-Research-Study/Accenture-Banking-Global-Distribution-Marketing-Consumer-Study.pdfla=en. (Accessed 23rd of August, 2020)

Accenture. (2019). *Global financial services consumer study.* https://www.accenture.com/_acnmedia/PDF-95/Accenture-2019-Global-Financial-Services-Consumer-Study.pdf. (Accessed 29th of August, 2020)

Acquisti, A., & Varian, H. R. (2005). Conditioning prices on purchase history. *Marketing Science*, *24*(3), 367–381.

Adam, M., Wessel, M., Benlian, A., et al. (2020). Ai-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, *9*(2), 204.

Ajenaghughrure, I. B., da Costa Sousa, S. C., & Lamas, D. (2020). Risk and trust in artificial intelligence technologies: A case study of autonomous vehicles. In *2020 13th international conference on human system interaction (hsi)* (pp. 118–123).

Ames, D. L., & Fiske, S. T. (2013). Intentional harms are worse, even when they're not. *Psychological science*, *24*(9), 1755–1762.

Amir, D., Jordan, M. R., & Rand, D. G. (2018). An uncertainty management perspective on long-run impacts of adversity: The influence of childhood socioeconomic status on risk, time, and social preferences. *Journal of Experimental Social Psychology*, *79*, 217–226.

Arkes, H. R., Dawes, R. M., & Christensen, C. (1986). Factors influencing the use of a decision rule in a probabilistic task. *Organizational behavior and human decision processes*, *37*(1), 93–110.

Asparouhova, E. N., Bossaerts, P., Rotaru, K., Wang, T., Yadav, N., & Yang, W. (2020). Humans in charge of trading robots: The first experiment. *Available at SSRN 3569435*.

Asparouhova, E. N., Cai, X., Meloso, D., Nielsen, J., Parlour, C. A., & Yang, W. (2022). For better or for worse: Algorithmic choice in experimental markets. *Available at SSRN*.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., ... Rahwan, I. (2018). The moral machine experiment. *Nature*, *563*(7729), 59–64.

Baker, E. H. (2014). Socioeconomic status, definition. *The Wiley Blackwell encyclopedia of health, illness, behavior, and society*, 2210–2214.

Banerjee, A. V., & Mullainathan, S. (2008). Limited attention and income distribution. *American Economic Review*, *98*(2), 489–93.

Bankins, S., Formosa, P., Griep, Y., & Richards, D. (2022). Ai decision making with dignity? contrasting workers' justice perceptions of human and ai decision making in a human resource management context. *Information Systems Frontiers*, 1–19.

Barboni, G., Cassar, A., Demont, T., et al. (2017). Financial exclusion in developed countries: a field experiment among migrants and low-income people in italy. *Journal of Behavioral Economics for Policy*, *1*(2), 39–49.

Baron, J. (2000). *Thinking and deciding*. Cambridge University Press.

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of personality and social psychology*, *54*(4), 569.

Bartling, B., & Fischbacher, U. (2012). Shifting the blame: On delegation and responsibility. *The Review of Economic Studies*, *79*(1), 67–87.

Batteux, E., Ferguson, E., & Tunney, R. J. (2019). Do our risk preferences change when we make decisions for others? a meta-analysis of self-other differences in decisions involving risk. *PloS one*, *14*(5), e0216566.

Bazerman, M. H., & Gino, F. (2012). Behavioral ethics: Toward a deeper understanding of moral judgment and dishonesty. *Annual Review of Law and Social Science*.

Belleflamme, P., & Vergote, W. (2016). Monopoly price discrimination and privacy: The hidden cost of hiding. *Economics Letters*, *149*, 141–144.

Bergemann, D., Brooks, B., & Morris, S. (2015). The limits of price discrimination. *American Economic Review*, *105*(3), 921–57.

Blalock, G., Just, D. R., & Simon, D. H. (2007). Hitting the jackpot or hitting the skids: Entertainment, poverty, and the demand for state lotteries. *American Journal of Economics and Sociology*, *66*(3), 545–570.

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.

Botta, M., & Wiedemann, K. (2020). To discriminate or not to discriminate? personalised pricing in online markets as exploitative abuse of dominance. *European journal of law and economics*, *50*(3), 381–404.

Boukherouaa, E. B., AlAjmi, K., Deodoro, J., Farias, A., & Ravikumar, R. (2021). Powering the digital economy: Opportunities and risks of artificial intelligence in finance. *Departmental Papers*, *2021*(024).

Bovens, L. (2009). The ethics of nudge. In *Preference change* (pp. 207–219).

Springer.

Bratman, M. (1987). Intention, plans, and practical reason. *Harvard Univ. Press*.

Brenner, L., & Meyll, T. (2020). Robo-advisors: a substitute for human financial advice? *Journal of Behavioral and Experimental Finance*, *25*, 100275.

Buiten, M. C. (2019). Towards intelligent regulation of artificial intelligence. *European Journal of Risk Regulation*, *10*(1), 41–59.

Camerer, C., & Weber, M. (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of risk and uncertainty*, *5*(4), 325–370.

Caplan, R. A., Posner, K. L., & Cheney, F. W. (1991). Effect of outcome on physician judgments of appropriateness of care. *Jama*, *265*(15), 1957–1960.

Capponi, A., Olafsson, S., & Zariphopoulou, T. (2022). Personalized robo-advising: Enhancing investment through client interaction. *Management Science*, *68*(4), 2485–2512.

Carpenter, J. (2015). Ibm's virginia rometty tells nu grads: Technology will enhance us. *Retrieved February*, *11*, 2019.

Carvalho, L. S., Meier, S., & Wang, S. W. (2016). Poverty and economic decision-making: Evidence from changes in financial resources at payday. *American economic review*, *106*(2), 260–84.

Casal, S., Ploner, M., & Sproten, A. N. (2019). Fostering the best execution regime: An experiment about pecuniary sanctions and accountability in fiduciary money management. *Economic Inquiry*, *57*(1), 600–616.

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, *56*(5), 809–825.

Cellan-Jones, R. (2014). Stephen hawking warns artificial intelligence could end mankind. *BBC news*, *2*(2014), 10–10.

Charness, G., Gneezy, U., & Henderson, A. (2018). Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization*, *149*, 74–87.

Checkland, P., & Holwell, S. (1998). *Information, systems, and information systems*. John Wiley & Sons Chichester.

Chen, D. L., Schonger, M., & Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97.

Cherkaoui, M., & Gilbert, S. J. (2017). Strategic use of reminders in an 'intention offloading'task: Do individuals with autism spectrum conditions compensate for memory difficulties? *Neuropsychologia*, *97*, 140–151.

Chugunova, M., & Sele, D. (2022). An interdisciplinary review of the experimental evidence on how humans interact with machines. *Journal of Behavioral and Experimental Economics*, 101897.

Cialdini, R. B., & Cialdini, R. B. (2007). *Influence: The psychology of persuasion* (Vol. 55). Collins New York.

Clark, A., & Chalmers, D. (1998). The extended mind. *analysis*, *58*(1), 7–19.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*(1), 67–80.

Davenport, T. H. (2019). *Can we solve al's' trust problem'?* (Vol. 60) (No. 2). SLOAN MANAGEMENT REVIEW ASSOC, MIT SLOAN SCHOOL MANAGEMENT 77 MASSACHUSETTS ....

Davenport, T. H., & Kirby, J. (2016). Just how smart are smart machines? *MIT Sloan Management Review*, *57*(3), 21.

De Lannoy, A., Leibbrandt, M., & Frame, E. (2015). A focus on youth: An opportunity to disrupt the intergenerational transmission of poverty.

South African child gauge, *2015*, 22–33.

Diamond, A. (2013). Executive functions. *Annual review of psychology*, *64*, 135–168.

Dietvorst, B. J., & Bharti, S. (2020). People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error. *Psychological science*, *31*(10), 1302–1314.

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.

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.

Dowling, K., Guhl, D., Klapper, D., Spann, M., Stich, L., & Yegoryan, N. (2020). Behavioral biases in marketing. *Journal of the Academy of Marketing Science*, *48*(3), 449–477.

Drewnowski, A., & Specter, S. E. (2004). Poverty and obesity: the role of energy density and energy costs. *The American journal of clinical nutrition*, *79*(1), 6–16.

Dubé, J.-P., & Misra, S. (2017). Personalized pricing and consumer welfare.

Duncan, G. J., Ziol-Guest, K. M., & Kalil, A. (2010). Early-childhood poverty and adult attainment, behavior, and health. *Child development*, *81*(1), 306–325.

Duval, E. R., Garfinkel, S. N., Swain, J. E., Evans, G. W., Blackburn, E. K., Angstadt, M., ... Liberzon, I. (2017). Childhood poverty is associated with altered hippocampal function and visuospatial memory in adulthood. *Developmental cognitive neuroscience*, *23*, 39–44.

D'Acunto, F., Prabhala, N., & Rossi, A. G. (2019). The promises and pitfalls of robo-advising. *The Review of Financial Studies*, *32*(5), 1983–2020.

D'Acunto, F., & Rossi, A. G. (2021). *Robo-advising*. Springer.

Fanning, J., Mullen, S. P., McAuley, E., et al. (2012). Increasing physical activity with mobile devices: a meta-analysis. *Journal of medical Internet research*, *14*(6), e2171.

Feier, T., Gogoll, J., & Uhl, M. (2022). Hiding behind machines: artificial agents may help to evade punishment. *Science and Engineering Ethics*, *28*(2), 1–19.

Fenech, E. P., Drews, F. A., & Bakdash, J. Z. (2010). The effects of acoustic turn-by-turn navigation on wayfinding. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 54, pp. 1926–1930).

Fenton-O'Creevy, M., Soane, E., Nicholson, N., & Willman, P. (2011). Thinking, feeling and deciding: The influence of emotions on the decision making and performance of traders. *Journal of Organizational Behavior*, *32*(8), 1044–1061.

Fildes, R., & Goodwin, P. (2007). Against your better judgment? how organizations can improve their use of management judgment in forecasting. *Interfaces*, *37*(6), 570–576.

Filiz, I., Judek, J. R., Lorenz, M., & Spiwoks, M. (2021). Reducing algorithm aversion through experience. *Journal of Behavioral and Experimental Finance*, *31*, 100524.

Filiz, I., Judek, J. R., Lorenz, M., & Spiwoks, M. (2022). Algorithm aversion as an obstacle in the establishment of robo advisors. *Journal of Risk and Financial Management*, *15*(8), 353.

Fischhoff, B. (1975). Hindsight is not equal to foresight: the effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human perception and performance*, *1*(3), 288.

Fisher, S. C., Hull, C., & Holtz, P. (1956). Past experience and perception:

memory color. *The American Journal of Psychology*, *69*(4), 546–560.

Flavián, C., Guinaliu, M., & Torres, E. (2006). How bricks-and-mortar attributes affect online banking adoption. *International Journal of Bank Marketing*.

Formosa, P., Rogers, W., Griep, Y., Bankins, S., & Richards, D. (2022). Medical ai and human dignity: Contrasting perceptions of human and artificially intelligent (ai) decision making in diagnostic and medical resource allocation contexts. *Computers in Human Behavior*, *133*, 107296.

Frank, D.-A., Chrysochou, P., Mitkidis, P., & Ariely, D. (2019). Human decision-making biases in the moral dilemmas of autonomous vehicles. *Scientific reports*, *9*(1), 1–19.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological forecasting and social change*, *114*, 254–280.

Furlough, C., Stokes, T., & Gillan, D. J. (2021). Attributing blame to robots: I. the influence of robot autonomy. *Human Factors*, *63*(4), 592–602.

Gandy, K., King, K., Streeter Hurle, P., Bustin, C., & Glazebrook, K. (2016). *Poverty and decision-making: How behavioural science can improve opportunity in the uk, the behavioural insights team (bit).*

Germann, M., & Merkle, C. (2022). Algorithm aversion in delegated investing. *Journal of Business Economics*, 1–37.

Germann, M., & Weber, M. (2018). Outcome bias in financial decision making. *Available at SSRN 3247148*.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., ... Rapoport, J. L. (1999). Brain development during childhood and adolescence: a longitudinal mri study. *Nature neuroscience*, *2*(10), 861–863.

Gilbert, N., & Stoneman, P. (2015). *Researching social life*. Sage.

Gilbert, S. J. (2015). Strategic use of reminders: Influence of both domain-general and task-specific metacognitive confidence, independent of objective memory ability. *Consciousness and Cognition*, *33*, 245–260.

Gino, F., Moore, D., & Bazerman, M. (2009). No harm, no foul: The outcome bias in ethical judgments. harvard business school nom working paper (08-080). *Available at SSRN 1099464*.

Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *The quarterly journal of economics*, *112*(2), 631–645.

Gnewuch, U., Morana, S., & Maedche, A. (2017). Towards designing cooperative and social conversational agents for customer service. In *Icis*.

Gomber, P., Kauffman, R. J., Parker, C., & Weber, B. W. (2018). On the fintech revolution: Interpreting the forces of innovation, disruption, and transformation in financial services. *Journal of Management Information Systems*, *35*(1), 220–265.

Gray, W. D., & Fu, W.-T. (2004). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive science*, *28*(3), 359–382.

Gray, W. D., Sims, C. R., Fu, W.-T., & Schoelles, M. J. (2006). The soft constraints hypothesis: a rational analysis approach to resource allocation for interactive behavior. *Psychological review*, *113*(3), 461.

Griskevicius, V., Ackerman, J. M., Cantú, S. M., Delton, A. W., Robertson, T. E., Simpson, J. A., … Tybur, J. M. (2013). When the economy falters, do people spend or save? responses to resource scarcity depend on childhood environments. *Psychological science*, *24*(2), 197–205.

Griskevicius, V., Tybur, J. M., Delton, A. W., & Robertson, T. E. (2011).

The influence of mortality and socioeconomic status on risk and delayed rewards: a life history theory approach. *Journal of personality and social psychology*, *100*(6), 1015.

Hackman, D. A., & Farah, M. J. (2009). Socioeconomic status and the developing brain. *Trends in cognitive sciences*, *13*(2), 65–73.

Hackman, D. A., Farah, M. J., & Meaney, M. J. (2010). Socioeconomic status and the brain: mechanistic insights from human and animal research. *Nature reviews neuroscience*, *11*(9), 651–659.

Hair, N. L., Hanson, J. L., Wolfe, B. L., & Pollak, S. D. (2015). Association of child poverty, brain development, and academic achievement. *JAMA pediatrics*, *169*(9), 822–829.

Hall, C. C., Zhao, J., & Shafir, E. (2014). Self-affirmation among the poor: Cognitive and behavioral implications. *Psychological science*, *25*(2), 619–625.

Hansen, P. G., et al. (2019). Nudging: To know'what works' you need to know why it works. *Journal of Behavioral Economics for Policy*, *3*(S), 9–11.

Hartfree, Y., & Collard, S. (2014). Poverty, debt and credit: An expert-led review. *University of Bristol*.

Harvey, C. R., Rattray, S., Sinclair, A., & Van Hemert, O. (2017). Man vs. machine: Comparing discretionary andsystematic hedge fund performance. *The Journal of Portfolio Management*, *43*(4), 55–69.

Haushofer, J., & Fehr, E. (2014). On the psychology of poverty. *science*, *344*(6186), 862–867.

Haushofer, J., & Fehr, E. (2019). Negative income shocks increase discount rates.

Haushofer, J., & Shapiro, J. (2013). Household response to income changes: Evidence from an unconditional cash transfer program in kenya. *Massachusetts Institute of Technology*, *24*(5), 1–57.

Heilman, R. M., Crişan, L. G., Houser, D., Miclea, M., & Miu, A. C. (2010). Emotion regulation and decision making under risk and uncertainty. *Emotion*, *10*(2), 257.

Helbing, D. (2019). Societal, economic, ethical and legal challenges of the digital revolution: from big data to deep learning, artificial intelligence, and manipulative technologies. In *Towards digital enlightenment* (pp. 47–72). Springer.

Hern, A. (2014). Okcupid: we experiment on users. everyone does. *The Guardian*.

Hidalgo, C. A., Orghian, D., Canals, J. A., De Almeida, F., & Martin, N. (2021). *How humans judge machines*. MIT Press.

Hill, A. (2015). Does delegation undermine accountability? experimental evidence on the relationship between blame shifting and control. *Journal of Empirical Legal Studies*, *12*(2), 311–339.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors*, *57*(3), 407–434.

Hohenstein, J., & Jung, M. (2020). Ai as a moral crumple zone: The effects of ai-mediated communication on attribution and trust. *Computers in Human Behavior*, *106*, 106190.

Holzmeister, F., Holmén, M., Kirchler, M., Stefan, M., & Wengström, E. (2022). Delegation decisions in finance. *Management Science*.

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-ai symbiosis in organizational decision making. *Business Horizons*, *61*(4), 577–586.

Jerčić, P., Astor, P. J., Adam, M., Hilborn, O., Schaff, K., Lindley, C., ... Eriksson, J. (2012). A serious game using physiological interfaces for emotion regulation training in the context of financial decision-making. In *20th european conference on information systems (ecis*

2012), barcelona (pp. 1–14).

Kahneman, D., & Tversky, A. (2013). Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part i* (pp. 99–127). World Scientific.

Kaplan, A., & Haenlein, M. (2019). Siri, siri, in my hand: Who's the fairest in the land? on the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, *62*(1), 15–25.

Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2016). Getting to the top of mind: How reminders increase saving. *Management Science*, *62*(12), 3393–3411.

Kashyap, P. (2018). *Machine learning for decision makers: Cognitive computing fundamentals for better decision making.* Apress.

Kaya, O., Schildbach, J., AG, D. B., & Schneider, S. (2019). Artificial intelligence in banking. *Artificial intelligence*.

Khan, S., Murray, R. P., & Barnes, G. E. (2002). A structural equation model of the effect of poverty and unemployment on alcohol abuse. *Addictive behaviors*, *27*(3), 405–423.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive science*, *18*(4), 513–549.

König-Kersting, C., Pollmann, M., Potters, J., & Trautmann, S. T. (2021). Good decision vs. good results: Outcome bias in the evaluation of financial agents. *Theory and Decision*, *90*(1), 31–61.

Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: a meta-analysis. *Journal of applied psychology*, *98*(6), 1060.

Lefgren, L., Platt, B., & Price, J. (2015). Sticking with what (barely) worked: A test of outcome bias. *Management Science*, *61*(5), 1121–1136.

Leib, M., Köbis, N. C., Rilke, R. M., Hagens, M., & Irlenbusch, B. (2021). The corruptive force of ai-generated advice. *arXiv preprint arXiv:2102.07536*.

Leonard, T. C. (2008). *Richard h. thaler, cass r. sunstein, nudge: Improving decisions about health, wealth, and happiness.* Springer.

Lesner, R. V. (2018). The long-term effect of childhood poverty. *Journal of Population Economics*, *31*(3), 969–1004.

Liang, G., Sloane, J. F., Donkin, C., & Newell, B. R. (2022). Adapting to the algorithm: how accuracy comparisons promote the use of a decision aid. *Cognitive research: principles and implications*, *7*(1), 1–21.

Lipina, S. J., & Posner, M. I. (2012). The impact of poverty on the development of brain networks. *Frontiers in human neuroscience*, *6*, 238.

Lipshitz, R. (1989). "either a medal or a corporal": The effects of success and failure on the evaluation of decision making and decision makers. *Organizational Behavior and Human Decision Processes*, *44*(3), 380–395.

Livingstone, S., & Helsper, E. (2007). Gradations in digital inclusion: Children, young people and the digital divide. *New media & society*, *9*(4), 671–696.

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *American economic review*, *90*(2), 426–432.

Loewenstein, G., Nagin, D., & Paternoster, R. (1997). The effect of sexual arousal on expectations of sexual forcefulness. *Journal of Research in Crime and Delinquency*, *34*(4), 443–473.

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*, 90–103.

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, *46*(4), 629–650.

Luarn, P., & Lin, H.-H. (2005). Toward an understanding of the behavioral intention to use mobile banking. *Computers in human behavior*, *21*(6), 873–891.

Lui, A., & Lamb, G. W. (2018). Artificial intelligence and augmented intelligence collaboration: regaining trust and confidence in the financial sector. *Information & Communications Technology Law*, *27*(3), 267–283.

Lyons, E. J., Lewis, Z. H., Mayrsohn, B. G., & Rowland, J. L. (2014). Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. *Journal of medical Internet research*, *16*(8), e3469.

Maeda, N. (2012). External working memory and the amount of distributed cognition. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 34).

Mahmud, H., Islam, A. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? a systematic literature review on algorithm aversion. *Technological Forecasting and Social Change*, *175*, 121390.

Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *science*, *341*(6149), 976–980.

Mayson, S. G. (2018). Bias in, bias out. *Bias Out (September 28, 2018)*, *128*.

McCarthy, N. (2016). *How much does poverty cost the uk every year?* `https://www.statista.com/chart/5428/how-much-does-poverty-cost-the-uk-every-year/`. (Accessed 26th of August, 2020)

McLaughlin, M., & Rank, M. R. (2018). Estimating the economic cost of childhood poverty in the united states. *Social Work Research*, *42*(2), 73–83.

Miller, A. P. (2018). Want less-biased decisions? use algorithms. *Harvard business review*, *26*.

Morgan, P. L., Patrick, J., Waldron, S. M., King, S. L., & Patrick, T. (2009). Improving memory after interruption: Exploiting soft constraints and manipulating information access cost. *Journal of Experimental Psychology: Applied*, *15*(4), 291.

Mousavi, S., Kheirandish, R., et al. (2017). Policy making with behavioral insight. *Journal of Behavioral Economics for Policy*, *1*(1), 41–46.

Mrkva, K., Posner, N. A., Reeck, C., & Johnson, E. J. (2021). Do nudges reduce disparities? choice architecture compensates for low consumer knowledge. *Journal of Marketing*, *85*(4), 67–84.

Mukherjee, A., & Nath, P. (2003). A model of trust in online relationship banking. *International journal of bank marketing*.

Mullainathan, S., & Shafir, E. (2014). *Scarcity: The true cost of not having enough*. Penguin books.

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological bulletin*, *126*(2), 247.

Nielsen, A. S. E., Sand, H., Sørensen, P., Knutsson, M., Martinsson, P., Persson, E., & Wollbrant, C. (2017). *Nudging and pro-environmental behaviour*. Nordisk Ministerråd.

Niszczota, P., & Kaszás, D. (2020). Robo-investment aversion. *Plos one*, *15*(9), e0239277.

OECD. (2019). *Governments must act to help struggling middle class.* `https://www.oecd.org/newsroom/governments-must-act-to-help-struggling-middle-class.htm`. (Accessed 26th of August, 2020)

Oexl, R., & Grossman, Z. J. (2013). Shifting the blame to a powerless intermediary. *Experimental Economics*, *16*(3), 306–312.

Oshri, A., Hallowell, E., Liu, S., MacKillop, J., Galvan, A., Kogan, S. M., & Sweet, L. H. (2019). Socioeconomic hardship and delayed reward discounting: Associations with working memory and emotional reactivity. *Developmental cognitive neuroscience*, *37*, 100642.

Osiurak, F., Navarro, J., Reynaud, E., & Thomas, G. (2018). Tools don't—and won't—make the man: A cognitive look at the future. *Journal of Experimental Psychology: General*, *147*(5), 782.

Óskarsdóttir, M., Bravo, C., Sarraute, C., Vanthienen, J., & Baesens, B. (2019). The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics. *Applied Soft Computing*, *74*, 26–39.

Ozer, E. J., Fernald, L. C., Weber, A., Flynn, E. P., & VanderWeele, T. J. (2011). Does alleviating poverty affect mothers' depressive symptoms? a quasi-experimental investigation of mexico's oportunidades programme. *International journal of epidemiology*, *40*(6), 1565–1576.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human factors*, *39*(2), 230–253.

Pariser, E. (2011). *The filter bubble: What the internet is hiding from you.* Penguin UK.

Parker, J. R., & Lehmann, D. R. (2011). When shelf-based scarcity impacts consumer preferences. *Journal of Retailing*, *87*(2), 142–155.

Pashchenko, I., Vu, D.-L., & Massacci, F. (2020). A qualitative study of dependency management and its security implications. In *Proceedings of the 2020 acm sigsac conference on computer and communications security* (pp. 1513–1531).

Pasquinelli, M. (2015). *Alleys of your mind: augmented intelligence and its traumas.* meson press.

Pelster, M., & Breitmayer, B. (2019). Attracting attention from peers: Excitement in social trading. *Journal of Economic Behavior & Organization*, *161*, 158–179.

Perkins, L., Miller, J. E., Hashemi, A., & Burns, G. (2010). Designing for human-centered systems: Situational risk as a factor of trust in automation. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 54, pp. 2130–2134).

Porter, A. (2017). Bioethics and transhumanism. In *The journal of medicine and philosophy: a forum for bioethics and philosophy of medicine* (Vol. 42, pp. 237–260).

Ragnedda, M., & Muschert, G. W. (2013). *The digital divide.* Routledge Florence, KY.

Ratner, R. K., & Herbst, K. C. (2005). When good decisions have bad outcomes: The impact of affect on switching behavior. *Organizational Behavior and Human Decision Processes*, *96*(1), 23–37.

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in cognitive sciences*, *20*(9), 676–688.

Risko, E. F., Medimorec, S., Chisholm, J., & Kingstone, A. (2014). Rotating with rotated text: A natural behavior approach to investigating cognitive offloading. *Cognitive Science*, *38*(3), 537–564.

Rossi, A. G., & Utkus, S. P. (2020). Who benefits from robo-advising? evidence from machine learning. *Evidence from Machine Learning (March 10, 2020).*

Rotchanakitumnuai, S., & Speece, M. (2003). Barriers to internet banking adoption: a qualitative study among corporate customers in thailand. *International Journal of Bank Marketing*.

Sacchi, S., & Cherubini, P. (2004). The effect of outcome information on doctors' evaluations of their own diagnostic decisions. *Medical education*, *38*(10), 1028–1034.

Sanders, N. R., & Manrodt, K. B. (2003). The efficacy of using judgmental versus quantitative forecasting methods in practice. *Omega*, *31*(6), 511–522.

Sargent, G. (2013). *Inequality is 'the defining issue of our time'.* https://www.washingtonpost.com/blogs/plum-line/wp/2013/12/04/inequality-is-the-defining-issue-of-our-time/. (Accessed 26th of August, 2020)

Sarsour, K., Sheridan, M., Jutte, D., Nuru-Jeter, A., Hinshaw, S., & Boyce, W. T. (2011). Family socioeconomic status and child executive functions: The roles of language, home environment, and single parenthood. *Journal of the International Neuropsychological Society*, *17*(1), 120–132.

Satterfield, K., Baldwin, C., de Visser, E., & Shaw, T. (2017). The influence of risky conditions in trust in autonomous systems. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 61, pp. 324–328).

Savage, L. J. (1972). *The foundations of statistics.* Courier Corporation.

Schmidt, U., Neyse, L., & Aleknonyte, M. (2019). Income inequality and risk taking: the impact of social comparison information. *Theory and Decision*, *87*(3), 283–297.

Seo, M.-G., & Barrett, L. F. (2007). Being emotional during decision making—good or bad? an empirical investigation. *Academy of Management Journal*, *50*(4), 923–940.

Shah, A. K., Mullainathan, S., & Shafir, E. (2012). Some consequences of having too little. *Science*, *338*(6107), 682–685.

Shah, A. K., Zhao, J., Mullainathan, S., & Shafir, E. (2018). Money in the mental lives of the poor. *Social Cognition*, *36*(1), 4–19.

Sharma, M. (2019). Augmented intelligence: a way for helping universities to make smarter decisions. In *Emerging trends in expert applications and security* (pp. 89–95). Springer.

Sheehy-Skeffington, J. (2020). The effects of low socioeconomic status on decision-making processes. *Current opinion in psychology*, *33*, 183–188.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703.

Silberg, J., & Manyika, J. (2019). Notes from the ai frontier: Tackling bias in ai (and in humans). *McKinsey Global Institute*, 1–6.

Slapø, H. B., & Karevold, K. I. (2019). Simple eco-labels to nudge customers toward the most environmentally friendly warm dishes: An empirical study in a cafeteria setting. *Frontiers in Sustainable Food Systems*, *3*, 40.

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *science*, *333*(6043), 776–778.

Spears, D. (2011). Economic decision-making in poverty depletes behavioral control. *The BE Journal of Economic Analysis & Policy*, *11*(1).

Sunstein, C. R. (2015). Nudging and choice architecture: Ethical considerations. *Yale Journal on Regulation, Forthcoming*.

Sunstein, C. R. (2016). Do people like nudges. *Admin. L. Rev.*, *68*, 177.

Sunstein, C. R. (2018, March). The rise of behavioral economics: Richard Thaler's misbehaving. *Journal of Behavioral Economics for Policy*, *2*(1), 53-57.

Taylor, R. K. (2000). Marketing strategies: Gaining a competitive advantage through the use of emotion. *Competitiveness Review: An International Business Journal*.

Thaler, R. H. (2000). From homo economicus to homo sapiens. *Journal of* *economic perspectives*, *14*(1), 133–141.

Thaler, R. H. (2018). *Nudge, not sludge* (Vol. 361) (No. 6401). American Association for the Advancement of Science.

Thaler, R. H., & Benartzi, S. (2004). Save more tomorrow™: Using behavioral economics to increase employee saving. *Journal of political Economy*, *112*(S1), S164–S187.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness.* Penguin.

Tomm, B. M., & Zhao, J. (2016). Scarcity captures attention and induces neglect: Eyetracking and behavioral evidence. In *Cogsci.*

Turkheimer, E., Haley, A., Waldron, M., d'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of iq in young children. *Psychological science*, *14*(6), 623–628.

Vrieze, S. I., & Grove, W. M. (2009). Survey on the use of clinical and mechanical prediction methods in clinical psychology. *Professional Psychology: Research and Practice*, *40*(5), 525.

Wagmiller, R. L., & Adelman, R. M. (2009). Childhood and intergenerational poverty: The long-term consequences of growing up poor.

Walsh, M. M., & Anderson, J. R. (2009). The strategic nature of changing your mind. *Cognitive psychology*, *58*(3), 416–440.

Ward, A. F. (2013). Supernormal: How the internet is changing our memories and our minds. *Psychological Inquiry*, *24*(4), 341–348.

Weis, P. P., & Wiese, E. (2019a). Problem solvers adjust cognitive offloading based on performance goals. *Cognitive Science*, *43*(12), e12802.

Weis, P. P., & Wiese, E. (2019b). Using tools to help us think: Actual but also believed reliability modulates cognitive offloading. *Human Factors*, *61*(2), 243–254.

West, A. (2007). Poverty and educational achievement: why do children from low-income families tend to do less well at school? *Benefits*, *15*(3), 283–297.

Wilkinson, T. M. (2013). Nudging and manipulation. *Political Studies*, *61*(2), 341–355.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin & review*, *9*(4), 625–636.

Wu, C. M., Schulz, E., Pleskac, T. J., & Speekenbrink, M. (2022). Time pressure changes how people explore and respond to uncertainty. *Scientific reports*, *12*(1), 1–14.

Yaniv, I. (2004). The benefit of additional opinions. *Current directions in psychological science*, *13*(2), 75–78.

Young, T. M., Martin, S. S., Young, M. E., & Ting, L. (2001). Internal poverty and teen pregnancy. *Adolescence*, *36*(142), 289.

Zhao, J., & Tomm, B. M. (2018). Psychological responses to scarcity. In *Oxford research encyclopedia of psychology.*