Essays on Sustainable Finance – The Role of Preferences, Incentives, and Information for Sustainable Investments

---

Dissertation

zur Erlangung des Grades eines Doktors der wirtschaftlichen Staatswissenschaften

(Dr. rer. pol.)

des Fachbereichs Rechts- und Wirtschaftswissenschaften

der Johannes Gutenberg-Universität Mainz

vorgelegt von

Fabian Braun, M.Sc.

in Mainz

im Jahre 2025

Erstgutachter: Prof. Dr. Andrej Gill

Zweitgutachterin: Prof. Dr. Christine Laudenbach

Drittgutachter: Prof. Dr. Kevin Bauer

Vorsitzender der Prüfungskommission: Prof. Dr. Florian Hett

Datum der mündlichen Prüfung: 24.01.2025

To my parents Dagmar and Ralf.

To my wife Madeleine.

Acknowledgments

Completing this PhD has been a journey filled with challenges, growth, and learning, and I am deeply grateful for the support and guidance I have received along the way.

First and foremost, I would like to express my heartfelt gratitude to my supervisor and co-author, Prof. Dr. Andrej Gill, for his invaluable mentorship, insightful feedback, and unwavering support over the past years. I am particularly grateful for our inspiring collaboration on three chapters of this dissertation. His guidance has been a true blessing, and I could not have wished for a better supervisor.

I am also profoundly grateful to my co-author, Prof. Dr. Florian Hett, who generously acted as a second supervisor, providing invaluable guidance, insightful feedback, and thoughtful input throughout my research. His support played a crucial role in the development of this work. I would like to extend my heartfelt thanks to the members of my thesis committee, Prof. Dr. Christine Laudenbach and Prof. Dr. Kevin Bauer, for their constructive input and valuable suggestions, which significantly enriched my work.

A heartfelt thanks goes to my colleagues and fellow researchers at Johannes Gutenberg University Mainz and the Leibniz Institute SAFE Frankfurt for their camaraderie, stimulating discussions, and collaboration. The supportive environment you have fostered made this journey not only productive but also enjoyable.

A special thanks goes to Lea, who carefully read my thesis, provided an abundance of insightful comments, and even printed the final version while I was hiking the mountains of Nepal. I am deeply thankful to Nina, whose invaluable guidance and counsel helped me navigate the first four years of my dissertation. I am equally grateful to my peer Max, who taught me so much about programming, greatly enriched my technical skills, and helped me bring my thesis across the finish line.

I would like to acknowledge IPP Mainz and the Leibniz Institute SAFE Frankfurt for providing the financial support necessary for this research.

On a personal note, I am forever indebted to my family and friends for their unwavering belief in me. To my parents, Dagmar and Ralf, for their endless love and encouragement, and to my wife Madeleine, for being my constant source of strength and unwavering support during the highs and lows of this journey. Her patience, understanding, and belief in me kept me grounded during the most challenging times. She celebrated my successes, lifted me during setbacks, and always reminded me of the bigger picture. I could not have made it without her.

Finally, to all those who have contributed to this work, whether directly or indirectly, thank you.

Abstract

This dissertation consists of four essays that address different aspects of the overall question of what drives sustainable investment decisions. They serve the purpose of analyzing the role of (ambiguous) information, examining the consequences of (motivated) beliefs, and investigating the effects of preferences. To answer the overarching question, three incentivized online experiments are conducted. The first essay deals with the role of preferences, incentives, and information for sustainable investment decisions and shows that motivated beliefs shape the acquisition of information, which is subsequently used for investment decisions. The second essay relates to the objective of this thesis by showing what determines investors' perceptions of what an appropriate definition of sustainable investments should entail. The paper provides causal evidence that investors' perspectives on this issue are influenced by motivated beliefs. These beliefs are shaped by investors' return expectations and their knowledge about the sustainability characteristics of their previously selected investments in the experiment. The third essay identifies the defining characteristics of the next generation of retail investors. Moreover, it examines the predictive power of behavioral, demographic, and portfolio characteristics in identifying sustainable investor types. Finally, the last essay analyzes the characteristics of wealthy private individuals with a propensity to invest sustainably. The results of this dissertation have implications for policymakers, researchers, and financial institutions.

Zusammenfassung

Diese Dissertation besteht aus vier Aufsätzen, die sich mit verschiedenen Aspekten der Frage beschäftigen, was nachhaltige Investitionsentscheidungen beeinflusst. Die Essays analysieren die Rolle von (mehrdeutigen) Informationen, untersuchen die Folgen von (motivierten) Überzeugungen und analysiere die Auswirkungen von Präferenzen. Zur Beantwortung der übergeordneten Fragestellung werden drei anreizkompatible Online-Experimente durchgeführt. Der erste Aufsatz analysiert die Rolle von Präferenzen, Anreizen und Informationen für nachhaltige Investitionsentscheidungen und zeigt, dass motivierte Überzeugungen die Auswahl von Informationen beeinflussen, welche im Anschluss für Investitionsentscheidungen genutzt werden. Der zweite Aufsatz knüpft an das Ziel dieser Dissertation an, indem er aufzeigt, was die Vorstellungen von Investoren darüber bestimmt, was eine angemessene Definition nachhaltiger Investments beinhalten sollte. Der Aufsatz liefert kausale Belege dafür, dass die Wahrnehmung der Anleger zu diesem Thema durch motivierte Überzeugungen beeinflusst wird. Diese Überzeugungen werden durch die Renditeerwartungen der Anleger und ihr Wissen über die Nachhaltigkeitseigenschaften der zuvor von ihnen im Experiment ausgewählten Anlagen geprägt. Der dritte Aufsatz identifiziert Charakteristiken zukünftiger Privatanleger und untersucht die Vorhersagekraft von Verhaltens-, demographischen und Portfoliomerkmalen bei der Identifizierung nachhaltiger Anlegertypen. Abschließend behandelt der letzte Aufsatz dieser Dissertation die Charakteristiken vermögender Privatpersonen mit einer Neigung zu nachhaltigen Investments. Die Ergebnisse der Dissertation sind für politische Entscheidungsträger, Forscher und Finanzinstitutionen von Bedeutung.

Contents

List of Tables

List of Figures

Chapter 1

1 Introduction

The climate crisis is one of society's greatest challenges. According to Cartry *et al.* (2023), it requires an estimated 7 trillion USD annually to finance the global green transition of the economy. However, current investments fall short, with less than 2 trillion USD being invested annually. In order to promote the greening of the economy, the European Commission has launched a series of initiatives. To increase transparency in the market for sustainable investments and ultimately attract more private capital for the financing of the transition, the EU requires companies to disclose the impact of their activities on the environment and society (CSRD), and define sustainable economic activities (EU Taxonomy). In addition, financial firms must disclose the effects of their financial products on society and the environment (SFDR), and investment advisors are required to consider sustainability preferences in the Suitability Assessment when advising financial products (MiFID II). To evaluate and improve the effectiveness of these regulations, it is necessary to understand what drives sustainable investment preferences. The rapidly growing literature has proposed several different explanations.

Existing research demonstrates that investors often have a preference to invest sustainably (Barreda-Tarrazona *et al.*, 2011; Hartzmark and Sussman, 2019; Bauer *et al.*, 2021). Nonetheless, what drives sustainable investments is still unresolved. Riedl and Smeets (2017); Bauer *et al.* (2021); Gutsche *et al.* (2023) posit that altruistic motives are the primary determinants of sustainable investments. Indeed, some investors are even willing to accept lower returns or higher fees in order to invest sustainably (Gutsche and Ziegler, 2019; Bauer *et al.*, 2021; Engler *et al.*, 2023; Heeb *et al.*, 2023). Conversely, Døskeland and Pedersen (2016) show that wealth framing is more effective than moral framing for information-seeking and sustainable investment behavior. In addition, Giglio *et al.* (2023) demonstrate that, independent of other investment motives, only those investors who expect sustainable investments to outperform conventional investments hold significant amounts of sustainable investments. It is important to note that the market for sustainable investments is characterized by ambiguity, as there is no universal definition of what constitutes a sustainable investment (Filippini *et al.*, 2024), and sustainability metrics often lack clarity (Dimson *et al.*, 2020; Billio *et al.*, 2021; Berg *et al.*, 2022). Additionally, the performance of sustainable stocks relative to conventional ones remains uncertain (Eccles *et al.*, 2014; Friede *et al.*, 2015; Pastor *et al.*, 2021; Avramov *et al.*, 2022; Bolton and Kacperczyk, 2023; Latino, 2023; Bolton and Kacperczyk, 2024). Such ambiguity provides fertile ground for motivated beliefs (Gino *et al.*, 2016), posing a potential challenge to investors' demand for sustainable investments. Given the initiatives to leverage financial markets for advancing the green transition (see, for example, Commission (2019a)), it is crucial to understand the factors that drive sustainable investments and the barriers that deter sustainable investment decisions. In this dissertation, I contribute four essays to this overall question. In these articles, my co-authors and I analyze the role of (ambiguous) information, examine the consequences of (motivated) beliefs, study the effects of preferences, and finally identify the characteristics of a typical sustainable investor.

In Chapter 2, which is joint work with Andrej Gill and Florian Hett, we contribute to the question of what drives sustainable investments by showing the effect of motivated beliefs on information acquisition and subsequent sustainable investment decisions. Moreover, we investigate the role of pecuniary and non-pecuniary motives in sustainable investments. To do so, we conduct an online experiment with various incentive and information treatments and observe investors' reactions to these treatments within an incentivized stock market game designed to elicit revealed preferences for socially responsible investments. This approach mitigates concerns about experimenter demand effects (Haaland *et al.*, 2023) and hypothetical bias (List and Shogren, 1998; List, 2001; Harrison, 2006). We find that participants in our sample exhibit a preference for sustainable investments and are willing to pay for sustainability information. Our results indicate that both moral considerations and (perceived) financial incentives play a role in motivating sustainable investments. In the second part of Chapter 2, we manipulate participants' return expectations and examine how this affects their information acquisition regarding the moral implications of sustainable investments and subsequent investment decisions. Our results demonstrate a skewed information acquisition, which hinders the efficient transmission of sustainability preferences. This skewed information acquisition not only rationalizes non-sustainable investment choices but also influences subsequent investment decisions based on the moral information acquired.

As previously elaborated, the definition of sustainable investments is still quite dispersed around the world. Moreover, even within the European approach, clients are offered several definitions of sustainability to suit their preferences (Commission, 2020a, 2022a). Chapter 3 is joint work with Andrej Gill and Florian Hett and contributes to the overarching question by examining the role of motivated beliefs in shaping perceptions of what sustainability definitions should entail. People's motivations influence how they gather and process information and recall past experiences, resulting in beliefs that seem objective but are actually biased (Gino *et al.*, 2016). When the context allows for a plausible justification to act selfishly while maintaining a moral self-image, individuals exploit these opportunities to prioritize self-interest over morality (Gino *et al.*, 2016; Dana *et al.*, 2007). We conduct an incentivized online experiment to investigate how investors' sustainability convictions are influenced by motivated beliefs. To introduce these motivated beliefs, we exogenously vary participants' return expectations and their knowledge about the sustainability attributes of their previously selected investments in the experiment. Moreover, we provide participants with three different definitions of sustainability and display pro and con arguments for each of the definitions. Our findings provide causal evidence that investors' perceptions of what a sustainability definition should entail are shaped by motivated beliefs, which emerge from the interplay between the sustainability attributes of their existing portfolio holdings within the experiment and return expectations. Through open-ended responses, we demonstrate that participants, despite being presented with identical arguments, apply and interpret these arguments differently to justify their decisions. We argue that policymakers should consider motivated beliefs in regulatory frameworks. For example, the MiFID II regulations should simplify the process of eliciting sustainability preferences to mitigate this self-serving behavior channel.

In Chapter 4, I contribute to the overall question by employing Machine Learning methodologies on experimental data from Chapter 2 to identify the characteristics of typical sustainable investors and evaluate the predictive power of behavioral traits, demographics, and portfolio attributes in classifying investor types. Recent research highlights various factors and demographic variables that may influence sustainable investing, such as risk and return expectations (Døskeland and Pedersen, 2016; Giglio *et al.*, 2021; Degryse *et al.*, 2023), altruistic tendencies (Riedl and Smeets, 2017; Bauer *et al.*, 2021), gender (Junkus and Berry, 2010; Anderson and Robinson, 2022; Gutsche *et al.*, 2023), age (Junkus and Berry, 2010; Riedl and Smeets, 2017; Haber *et al.*, 2022), educational background (Junkus and Berry, 2010; Anderson and Robinson, 2022; Montagnoli and Taylor, 2024), income and wealth (Haber *et al.*, 2022; Gutsche *et al.*, 2023), and financial literacy (Anderson and Robinson, 2022; Gutsche *et al.*, 2023; Montagnoli and Taylor, 2024; Auzepy *et al.*, 2024; Filippini *et al.*, 2024). Nevertheless, the extent to which these factors influence sustainable investing remains inconclusive, with some studies indicating minimal or no significant effects (Riedl and Smeets, 2017; Haber *et al.*, 2022; Anderson and Robinson, 2022; Gutsche *et al.*, 2023). To contribute to the overall research question, I start by applying K-Means Clustering to construct data-driven investor personas (McGinn and Kotamraju, 2008; Miaskiewicz and Kozar, 2011; Akre *et al.*, 2019; Nielsen, 2019). The analysis results in two typical investors: a non-sustainable and a sustainable investor. Subsequently, I use Random Forests, XGBoost, K-Nearest Neighbor, and Logistic Regression models to forecast sustainable investor profiles. The analysis reveals that while individual attributes like financial literacy, education, and income provide limited predictive power, portfolio characteristics substantially improve the accuracy of the models.

Finally, in Chapter 5, in collaboration with Andrej Gill and Florian Hett, I address the research question of this thesis by examining the unique characteristics of sophisticated, affluent private investors who are inclined toward sustainable investing. This is especially relevant as individuals with a net worth between 100,000 USD and 1 million USD possess 39.4% of all private financial assets (Credit Suisse, 2023). Thus, due to their significant financial capacity, these individuals are central to financing the green transition. However, most research on sustainable finance tends to focus on demographic groups with less wealth. We implement an incentivized online experiment with wealthy clients from a German private bank, allowing us to measure their revealed preferences for sustainable investments. This data is then matched with the bank's Investor Suitability Assessment, which includes details on income, wealth, risk preferences, experience with financial products, age, and education. Our findings reveal a high level of consistency between the investors' self-reported information and the experimental data. We identify a clear preference for sustainable investments among affluent private investors and find a significant correlation between altruistic preferences and sustainable investments. For investors who are new to ESG, this correlation is mainly driven by impact altruism, while for those familiar with ESG, both impact and warm glow altruism influence their preferences. Additionally, these investors are willing to pay for information that helps them to make sustainable investment decisions. Finally, Chapter 5 shows that investors respond to information about the risk-adjusted performance of sustainable investments.

The remainder of this thesis is structured as follows: Chapter 2 analyzes what factors drive sustainable investments and studies the relationship between motivated beliefs, information acquisition, and subsequent investment decisions. The following chapter deepens the understanding of the role of motivated beliefs and describes their role in defining suitable sustainability definitions. Chapter 4 uses Machine Learning algorithms to understand who the sustainable investors are, while Chapter 5 analyzes the factors that influence wealthy private investors to invest sustainably. Last, Chapter 6 concludes.

Chapter 2

2 What Drives Socially Responsible Investments: The Role of Preferences, Incentives, and Information

Fabian Braun[a,b] Andrej Gill[a,c] Florian Hett[a,d]

August 2024

Abstract

In order to gain insights into the factors influencing socially responsible investments (SRI), we present the results of an incentivized online stock market game involving 643 participants. The study analyzes the role of preferences, incentives, and information and provides two key findings. First, moral and financial considerations play a role in sustainable investment decisions. We show that warm glow and impact altruists invest significantly more sustainably than non-altruists. Moreover, participants respond to information regarding the moral implications of sustainable investments, as well as to (perceived) financial incentives. Second, we demonstrate a behavioral mechanism that prevents efficient sustainable investments. Investors opportunistically acquire information on the moral consequences of SRI. As investors subsequently use the acquired information to make investment decisions, this skewed information acquisition prevents the efficient transmission of existing sustainability preferences. Our findings have implications for researchers trying to understand what drives sustainable investments and policymakers intending to attract private capital to finance the green transition of the economy.

Keywords: ESG, Sustainable Finance, Behavioural Finance, Stock Market Game, Experiment

[a]We gratefully acknowledge research support from the Leibniz Institute for Financial Research SAFE and the IPP Mainz. We thank Jonas Schwarze and Lea Voigt for outstanding research assistance and seminar participants at SAFE, Goethe University Frankfurt, and Johannes Gutenberg University Mainz for very helpful comments and discussions. The joint ethics committee of the Faculty of Economics and Business Administration of Goethe University Frankfurt and the Gutenberg School of Management & Economics of the Faculty of Law, Management and Economics of Johannes Gutenberg University Mainz granted the project exempt status. The experiment is pre-registered under AEARCTR-0008178 and if not stated otherwise, we follow a pre-registered analysis plan. The authors declare no conflict of interest.

[b]Johannes Gutenberg University Mainz; Jakob-Welder-Weg 9, 55128 Mainz, Germany. Phone: +49 6131 3927901. Email: braun@uni-mainz.de

[c]Johannes Gutenberg University Mainz; Jakob-Welder-Weg 9, 55128 Mainz, Germany. Phone: +49 6131 3922979. E-Mail: gill@uni-mainz.de

[d]Johannes Gutenberg University Mainz; Johann-Joachim-Becher-Weg 31, 55128 Mainz, Germany. Phone: +49 6131 39-20567. E-Mail: florian.hett@uni-mainz.de

2.1 Introduction

A significant proportion of investors has a preference for sustainable investments (Barreda-Tarrazona *et al.*, 2011; Hartzmark and Sussman, 2019; Bauer *et al.*, 2021). Indeed, some investors are even willing to accept lower returns or higher fees in order to invest sustainably (Gutsche and Ziegler, 2019; Bauer *et al.*, 2021; Engler *et al.*, 2023; Heeb *et al.*, 2023). However, the question of what drives sustainable investments remains unresolved. Riedl and Smeets (2017); Bauer *et al.* (2021); Gutsche *et al.* (2023) posit that altruistic motives are the primary drivers of sustainable investments. Conversely, Døskeland and Pedersen (2016); Giglio *et al.* (2023) argue that financial considerations are the main determinants. To date, there are no convincing guidelines that define and label sustainable investments consistently across the globe (Cordes *et al.*, 2024; Filippini *et al.*, 2024).[1] This may have significant implications for investors, as they can readily access information on a vast array of investment strategies, potentially impeding their ability to invest sustainably. As one example, Braun *et al.* (2024a) illustrate that motivated beliefs influence investors' convictions of what sustainability should entail in a setting with ambiguous definitions of sustainability. In this paper, we want to analyze what influences sustainable investments and whether frictions in the financial markets prevent an effective transmission of sustainability preferences.

To analyze these questions, we employ an incentivized online experiment that allows us to introduce several incentive and information treatments and observe participants' reactions to these treatments in a stock market game. This stock market game allows for the elicitation of revealed preferences for socially responsible investments (SRI) while maintaining complete control over the information available to participants. Study subjects allocate their money between real stocks, while we mask the true names of the stocks to prevent them from using outside information. In order to ensure incentive compatibility, we pay out one of the chosen portfolios to the investors after the experimental session and invest in the real stock market according to the choices of the individuals within the stock market game. To analyze the (perceived) costs of investing in sustainable stocks, we introduce an information and an incentive treatment. In the information treatment, investors' risk and return expectations are shifted by exogenously assigning them articles stating the risk-adjusted overperformance (underperformance) of sustainable stocks. In contrast, in the incentive treatment, participants receive a direct financial incentive (disincentive) to invest sustainably. Finally, we randomly assign 50% of the participants to choose between two research articles. One article describes the moral advantages of SRI, while the other describes the moral disadvantages. The remaining participants serve as the control group. The endogenous information acquisition reflects the moral flexibility of investors who

[1] We are aware of Europe's attempts to establish a clear and binding sustainability framework, which represents a step in the right direction. However, these are local regulations, while the capital markets are global. For further details and an analysis of potential shortcomings of these regulations, please see Dumrose *et al.* (2022); Cordes *et al.* (2024). Additionally, the International Financial Reporting Standards (IFRS) have attempted to establish a global benchmark for sustainability reporting, but the extent of uptake remains unclear (IFRS, 2024).

can find information to justify their (non-) SRI decisions, given the large amount of conflicting information in the market for sustainable investments. Finally, we employ incentivized dictator games to elicit altruistic preferences and assess demographic characteristics as well as risk and return expectations.

The experiment yields three main findings. First, participants value sustainability and are willing to pay around 6.75% of their endowment for ESG Score information assisting them to make sustainable investment decisions. This predilection for sustainability is associated with altruistic preferences. The results indicate that warm glow altruists and impact altruists hold more sustainable portfolios than non-altruists.[2] This has implications for regulators and financial institutions. Secondly, our findings demonstrate that investors respond to the perceived costs of sustainable investments. We show a significant impact of shifting the SRI risk and return beliefs on sustainable investments. For instance, exposure to ESG overperformance (underperformance) information increases (decreases) ESG Scores relative to their previous portfolio allocation by 33.6% (14.2%). In light of the political agenda to increase SRI,[3] a promising avenue could be to provide tax discounts on returns or to decrease the costs of investments by subsidizing this asset class in retirement plans. Third, we provide evidence that investors opportunistically acquire information on the moral consequences of SRI and then actively use this information for their subsequent investment decisions. To illustrate, suppose investors have a clear incentive to invest sustainably: They are 31% less likely to select information detailing the moral disadvantages of sustainable investments than investors with a clear incentive not to invest sustainably. This effect is driven by warm glow altruists who wish to feel good about their investment decisions. Investors not only selectively acquire information but also act on it. Appropriate matching indicates that reading information on the moral shortcomings of ESG decreases sustainable investments by 13.58 percentage points compared to reading information on the moral advantages of sustainable investments. The demonstrated behavioural bias results in an inadequate transmission of existing sustainability preferences in the financial markets and calls for regulations that resolve information frictions in the market for sustainable investments.

Our project adds to the literature in multiple ways. First, we contribute to the literature on sustainable investments. Investors demonstrate a preference for sustainability (Barreda-Tarrazona *et al.*, 2011; Hartzmark and Sussman, 2019; Pástor *et al.*, 2020; Bauer *et al.*, 2021) and are even willing to forgo returns and pay higher fees to invest sustainably (Gutsche and Ziegler, 2019; Heeb *et al.*, 2023; Baker *et al.*, 2022; Laudi *et al.*, 2023; Engler *et al.*, 2023). We experimentally confirm that investors value sustainability and add the direct measurement

[2]For a discussion on the different motivations behind charitable giving, please see Andreoni (1989, 1990). He introduces the concepts of pure altruism and warm glow, which we refer to as impact altruism and warm glow altruism. Warm glow altruists exhibit a selfish motivation for giving, donating to feel good about themselves. Conversely, impact altruists are completely selfless and give without expecting any personal reward. Impact altruists are solely motivated by the benefits received by the recipient of the donation.

[3]One illustrative example is the European Green Deal initiative, which aims for net zero emissions by 2050 and argues that the attraction of private capital is essential to achieve that goal (Commission, 2019a).

of investors' willingness to pay for ESG Score information helping them to make sustainable investment choices.

Moreover, the literature identifies altruistic preferences as one of the key drivers for sustainable investments (Riedl and Smeets, 2017; Bauer *et al.*, 2021; Heeb *et al.*, 2023; Gutsche *et al.*, 2023). The seminal work of Riedl and Smeets (2017) correlates social preferences and risk and return expectations with true sustainability holdings of Dutch bank clients and points out that non-pecuniary rather than financial motivations drive sustainable investments. In addition, Heeb *et al.* (2023) examine investors' willingness to pay for sustainable investments and demonstrate that this willingness to pay does not scale with more impact. They argue that warm glow rather than impact altruism motivates sustainable investments. To our knowledge, our study is the first to experimentally disentangle impact altruists and warm glow altruists, showing that both altruistic types correlate with sustainable investment behavior. The distinction between impact altruism, defined as utility derived from the benefits to others, and warm glow altruism, characterized by personal satisfaction from the act of doing good (Andreoni, 1989, 1990), is crucial for developing strategies that align with investor motivations and for regulatory frameworks that support genuine sustainability outcomes. Braun *et al.* (2024c) find in a similar setting with wealthy private investors a significant correlation with impact altruism, while the effect of warm glow altruism is mixed.

The second main factor for sustainable investments identified in the literature is risk and return expectations on SRI (Døskeland and Pedersen, 2016; Riedl and Smeets, 2017; Dong *et al.*, 2022; Gutsche *et al.*, 2023; Giglio *et al.*, 2023; Braun *et al.*, 2024c). Giglio *et al.* (2023) survey Vanguard clients and observe that, although investors generally expect sustainable investments to underperform the market, only those anticipating higher returns hold significant amounts of ESG assets. Most similar to our study is Døskeland and Pedersen (2016), who use a natural field experiment within an online banking context to demonstrate that financial framing is more effective than moral framing in increasing sustainable investments. Our approach differs in two ways: First, we investigate not only the effect of positive performance information but also analyze the effect of negative performance information on sustainable assets. Second, we introduce direct positive and negative financial incentives, which allow us to calculate the needed monetary incentive to increase sustainable investments by one percentage point. More broadly, we contribute to the discussion on whether individuals driven by social reasons decrease their intrinsic motivation to invest sustainably if their actions are perceived as financially advantageous (Frey and Oberholzer-Gee, 1997; Mellström and Johannesson, 2008; Gneezy *et al.*, 2011; Brodback *et al.*, 2019). Our findings demonstrate a significant reaction to increases and decreases in risk and return expectations, consistent with the explanation that increases (decreases) in financial utility outweigh decreases (increases) in social utility. Overall, our results contribute to the understanding that both moral and financial factors influence sustainable investment behavior.

Furthermore, our study adds to the literature on motivated beliefs[4] and information acquisition (Bénabou and Tirole, 2002; Dana *et al.*, 2007; Golman *et al.*, 2017; Sharot and Sunstein, 2020; Golman *et al.*, 2022; Eyting, 2022) by linking these concepts to ambiguous information in the context of sustainability. Individuals tend to reason their way to conclusions that favor them (Epley and Gilovich, 2016). The authors note that people's motivation influences how they gather and process arguments or recall past experiences, resulting in beliefs that appear objective but are actually biased. Individuals even prefer to have some "moral wiggle room", which allows them to engage in selfish behavior without taking responsibility for the harm of their choices to others (Dana *et al.*, 2007). When the context provides sufficient flexibility to allow plausible justifications for acting selfishly while remaining moral, people take advantage of such opportunities to prioritize self-interest at the expense of morality (Dana *et al.*, 2007; Gino *et al.*, 2016). Moreover, individuals stop seeking for information when they have an outcome that favors them (Chen and Heese, 2021; Golman *et al.*, 2022) or even actively avoid information to feel better about themselves (Grossman and van der Weele, 2017; Golman *et al.*, 2017; Momsen and Ohndorf, 2020). In the context of evaluating sustainability definitions, Braun *et al.* (2024a) show that the presence of different definitions of sustainability, each with ample justification for being considered sustainable, leads participants to choose the one that offers them the greatest benefit. Most similar to our study is Ambuehl (2024), which demonstrates that financial incentives can systematically bias how individuals obtain and process information about transactions that may be disadvantageous to themselves, such as consuming insects or taking risks. Our study differs from the literature mentioned above by demonstrating that the perception of financial incentives affects the acquisition of information regarding the moral implications of sustainable investments, which subsequently influences investment behavior.

The remainder of the paper is structured as follows: Section 2.2 provides details on the experimental design, the data collection, and key variables. Section 2.3 presents the main results, and Section 2.4 discusses the results and concludes.

2.2 Experimental Design, Data Collection, and Key Variables

At the heart of our experiment is an incentivized stock market game where participants are asked to assemble portfolios of real-world stocks under various information and incentive settings. The experiment comprises four stages. After each stage, we elicit investors' preferences for sustainable investments.[5] In the remainder of this section, we describe the experimental design and provide an overview of the data collection process.

[4]See Epley and Gilovich (2016) for an overview on motivated beliefs.

[5]We use a comparable stock market game in Braun *et al.* (2024c) for a sample of experienced wealthy private investors, which allows us to show distinctive characteristics of experienced investors. Detailed instructions on all stages and measures are available upon request.

Measurement of Revealed Sustainability Preferences: To elicit investors' revealed sustainability preferences, we employ an incentivized stock market game, designed to maintain precise control over the information available to participants and to emphasize actual decision-making. This approach effectively minimizes hypothetical bias (List and Shogren, 1998; List, 2001; Harrison, 2006; Harrison *et al.*, 2008) and reduces the potential for experimenter demand effects (Haaland *et al.*, 2023), as participants make real financial decisions with tangible results.

Account Balance:
50.000 Punkte

Name	Share Price (2019)	Past Performance (2018-2019)	P/E Ratio (2019)	Revenue (2020)	ESG Risk Scores (2020)
Firma 56	100,76	-17,56%	11,55	45.192.000.000	39,3
Firma 80	37,28	6,77%	29,39	31.303.440.000	25,1
Firma 7	12,92	4,78%	28,08	80.185.000.000	16,8
Firma 27	19,71	-15,00%	8,89	76.644.960.000	28,0
Firma 76	36,10	0,09%	8,06	3.160.200.000	7,8
Firma 94	19,02	6,07%	14,07	8.585.892.000	16,0

Figure 1: Trading desk

Notes: This figure presents the (translated) trading desk interface used by participants during the stock market game. Each participant is endowed with 50,000 points and can allocate this endowment among five different stocks selected from an assortment of 20. Initially, participants view key figures for each stock such as share price, past performance, price-earnings ratio, revenue, and ESG Risk Score. By clicking on the stock's name, participants receive additional information, as illustrated in Figure 5, and have the option to purchase the stock.

Each participant is provided with a trading account to purchase up to five stocks from an assortment of 20, representing a cross-section of the market.[6] Real share prices and attributes are used in the experiment, but company names are anonymized and prices are converted to an experimental currency (points). This ensures that market gains or losses are directly reflected in the experimental currency, simulating real-world stock behavior in a controlled environment. Figure 1 provides a screenshot of the trading interface used in the game. Hovering over key figures in the interface displays a short explanation of the indicator, and clicking on the stocks name opens a detailed fact sheet with ten key figures for that stock, enhancing transparency and aiding decision-making (see Figure 5 in the appendix for an example). Participants receive publicly available information on past performance, dividends, price-to-earnings ratios, volatility, and the stock price as of January 2019. Additionally, they are presented with 2020 data on equity value, number of employees, revenue, debt ratio, and ESG Risk Score. Each participant starts with a trading account containing 50,000 points to purchase up to five different stocks at January

[6]The selection process entails the application of several filters to a list of 100 publicly traded companies sourced from Yahoo Finance. (1) Actual companies, excluding funds; (2) Fiscal year ending in December; (3) Classification into ESG Risk Score ranges of 0-10, 10-20, 20-30, 30-40, 40+; and (4) Market cap categorizations as small/mid-cap, large-cap, or mega-cap. The final selection of 20 companies is made randomly to ensure proportional representation from each ESG category and market capitalization level. In the event that there are insufficient sustainable firms in the largest market capitalization bin, firms from the next smaller bin are used.

2019 prices from a pool of 20 stocks, using the buy button positioned above and below the fact sheet.[7] The stock market game is conducted after each of the first three stages explained below. This approach yields three distinct measures of sustainability for each portfolio allocation. The first, *ESG Score*, is the average ESG Score of the selected five stocks.[8] Higher *ESG Scores* imply greater sustainability. As this indicator provides the most comprehensive information on sustainability, it is the preferred measure in this paper. The second indicator, *# Sustainable*, counts the number of sustainable stocks in a portfolio.[9] The third measure, *# Sustainable d.*, is an indicator variable set to one if the portfolio includes more than two sustainable stocks. To mitigate concerns regarding survey fatigue (Haaland *et al.*, 2023), in the last part of the experiment, participants state their sustainability preferences by allocating their endowment between a sustainable and a non-sustainable portfolio with otherwise similar characteristics. The measure *% SRI* is the share allocated to the sustainable portfolio in each of the last five portfolio decisions.

Stage 1 - Measurement of Preferences, Expectations, and Demographics: The experimental literature identifies altruism, risk aversion, risk and return expectations on sustainable investments, gender, income, and financial literacy as possible determinants of SRI (Riedl and Smeets, 2017; Gutsche and Ziegler, 2019; Bauer *et al.*, 2021; Haber *et al.*, 2022; Gutsche *et al.*, 2023; Montagnoli and Taylor, 2024). To elicit incentive-compatible measures of altruism, we combine two dictator games paired with a charity where participants allocate eight euros between themselves and a selected charity (Crumpler and Grossman, 2008; Tonin and Vlassopoulos, 2010). The experimental design incorporates two types of dictator games: a standard dictator game and a "money-burning" dictator game. Unlike in the standard game, in the "money-burning" dictator game, the charity receives eight euros regardless of the participant's donation. This setup facilitates the identification of non-altruists and altruists and categorizes the latter in a second step according to their altruistic motivations into impact altruists and warm glow altruists. (i) *altruists* indicate individuals that donate something in at least one of the games and reflect a combination of warm glow and impact altruistic motivations; (ii) *impact altruist*s are individuals that donate exclusively in the standard dictator game. They derive utility solely from the benefit of the recipient of the donation and consider the giving of others as a perfect substitute (Andreoni, 1989, 1990); (iii) *warm glow altruist*s indicate individuals that donate in both dictator games. These individuals receive satisfaction from the act of giving itself — hence, giving by others is not a perfect substitute. Furthermore, we elicit risk aversion (Sutter *et al.*,

[7] Participants are required to select exactly five stocks, but they may choose to purchase the same stock multiple times. Using the prices as of January 2019 allows us to instantly calculate the one-year portfolio returns and pay out participants accordingly.

[8] In the experiment, participants are presented the raw ESG Risk Scores from Sustainalytics, where a higher score indicates lower sustainability. For clarity, we linearly transform the ESG Risk Scores into ESG Scores using the formula *ESG Score = (2*mean(ESG Risk Score) - ESG Risk Score)*.

[9] Sustainable stocks are characterized by an ESG Risk Score below 20 (color-coded in grey and yellow). As 40% of the stocks in the experiment are defined as sustainable, a random choice results on average in two sustainable stocks.

2013), risk and return expectations (Riedl and Smeets, 2017), financial literacy (Hastings *et al.*, 2013; Lusardi and Mitchell, 2008), ESG knowledge, and a set of demographics. See Table 7 in the appendix for a detailed definition of each of the variables. Stage 1 of the experiment concludes with investors participating in the stock market game as described above to get a baseline demand for sustainable investment.

Stage 2 - Measurement of the Willingness to Pay (WTP) for ESG Score Information: To elicit investors' consequential WTP for ESG Score information, we endow each participant with a trading account of 50,000 points, which they can use to buy ESG Scores and stocks in the subsequent steps. We rely on the Becker–DeGroot–Marschak method (Becker *et al.*, 1964) to determine investors' WTP for ESG Scores. If a participant's WTP is equal to or higher than a random price [0-10,000 points], they gain access to the ESG Scores and must pay the random price. If their WTP is lower than the random price, they neither receive the ESG Scores nor have to pay anything. The variable *WTP* is measured by the number of points an investor is willing to sacrifice to acquire the ESG Score information in the second portfolio decision. In the final step of Stage 2, investors participate in the stock market game with the difference to the baseline decision that, depending on their previous choices, they may not have access to ESG Scores.

Stage 3 - Information Treatment - Reaction to Performance Information: The scientific literature presents conflicting views on the financial implications of integrating sustainability factors into the investment process. Some theoretical and empirical studies argue in favor of higher risk-adjusted returns for non-sustainable assets (Hong and Kacperczyk, 2009; Chava, 2014; Pedersen *et al.*, 2021; Bolton and Kacperczyk, 2021; Pastor *et al.*, 2021; Avramov *et al.*, 2022; Bolton and Kacperczyk, 2024). In contrast, others suggest that, in the long run, the risk-adjusted performance of sustainable assets is better or at least not worse (Kempf and Osthoff, 2007; Eccles *et al.*, 2014; Dimson *et al.*, 2015; Friede *et al.*, 2015; Avramov *et al.*, 2022; Berg *et al.*, 2022). To investigate participants' causal reactions to changes in return expectations of sustainable stocks, we exogenously vary the information provided about the risk-adjusted performance of firms with high (low) sustainability ratings in an information treatment. We randomly assign 50% of participants to receive a management summary titled *"Why investing in stocks with low ESG Risk Scores makes sense from a performance perspective"*, while the remaining 50% receive a summary titled *"Why investing in stocks with low ESG Risk Scores makes no sense from a performance perspective"*. Both summaries commence with a brief executive summary of the subsequent pages, after which the factors environmental (E), social (S), and governance (G) are introduced in a neutral manner. Scientific papers are utilized to elucidate the impact of each factor on stock performance (further details are available upon request). Finally, in the concluding phase of Stage 3, investors again engage in the stock market game as described.

Stage 4 - Incentive Treatment - Reaction to Direct Financial Incentives in Information Acquisition and Sustainable Investments: In the final part of the experiment, we delve deeper into the role of direct financial incentives for information acquisition and subsequent investment decisions. The experimental design involves four parts and is inspired by Ambuehl (2024). First, participants take part in an incentive treatment and receive information specifying the magnitude of the financial incentive to invest sustainably in the subsequent portfolio allocation. They are randomly assigned to either the positive incentive group or the negative incentive group. Specifically, the proctor subsidizes (reduces) the amount of sustainable investments made by participants in the positive (negative) incentive group by 10%. Second, 50% of all participants are randomly selected to choose and read one of two research articles that emphasize, respectively, the moral upsides or downsides of investing sustainably. Figure 2 displays the (translated) decision screen. Participants select between the articles *"Why investing in stocks with a low ESG Risk Rating makes sense from a moral perspective"* and *"Why investing in stocks with a low ESG Risk Rating makes no sense from a moral perspective"* by clicking the respective button. The endogenous choice of the article reflects the investment decision-making process in the real world, whereby investors select the information they deem most relevant. The remaining 50% of the participants serve as the control group and skip this part of the study.

Figure 2: Selection of one of two articles on the moral implications of SRI
Notes: This figure displays investors' decision screen for selecting between two articles on the moral perspectives of investments in stocks with low ESG Risk Ratings (translated in English).

Third, all participants are provided with a trading account of 50,000 points and allocate their endowment between a sustainable and a non-sustainable portfolio by moving a slider. We dynamically display participants the value of the high-sustainability portfolio, the value of the low-sustainability portfolio, and the overall portfolio value in their incentive group at the time of their investment (see Figure 6 in the appendix for an example). Fourth, we ask participants about their portfolio decisions (the two portfolios are the same as before) for different magnitudes of the incentive (-20%, -10%, 0%, +10%, +20%). Individuals move a slider for each incentive level to allocate their endowment of 50,000 points between the two funds. The decision the individual made in the previous portfolio allocation cannot be changed and is shown for reference.

Data Collection, Payment of Participants, and Incentive Compatibility: Data collection took place in an oTree-programmed (Chen *et al.*, 2016) online experiment with participants of the subject pools from Goethe University Frankfurt and Johannes Gutenberg University Mainz between September 6th and 30th, 2021. To ensure incentive compatibility, we pay out one of the dictator games or the risk elicitation task from the first stage and the value of one of the portfolios from January 2020 in stages 1 to 4 at an exchange rate of 1 point = 0.0002 euros. Specifically, participants are informed that we will randomly select one task from Stage 1 (each task is drawn with a 33.33% probability) and one of the eight portfolios from stages 1 to 4 (each portfolio is selected with a 12.5% probability). If the multiple price list risk elicitation task is selected, we randomly determine the row to pay out (each of the ten rows are equally likely). Additionally, participants are informed that the randomly selected portfolio will be purchased on the real stock market, with 20% of its actual value financed through private funds. This ensures that the participants are held morally accountable for their investments. All portfolio holdings are held for a period of 12 months. Payments to participants and the selected charity were made after the completion of the experiment. The median time spent in the experiment is 50 minutes, and the median earnings are 18 euros per participant. This equates to an hourly wage of 21.60 euros, which is substantially more than students typically earned as student assistants in 2021 (10.91-12.68 euros). We use these comparably high incentives to ensure that the trades in the stock market game reflect the choices participants would make on the real stock market.

Summary Statistics: The final sample comprises 643 participants.[10] Table 19 provides summary statistics for all variables. The sample consists of 63% females, with an average monthly net income of approximately 1,000 euros, which is higher than the median German student who had a monthly income of below 500 euros in 2021 (Statista, 2021). The average age of participants is 24 years old, with 46% of them holding bonds, stocks, or funds. This is a relatively high proportion, as only 15% of Germans aged between 20 and 29 years invest in the stock market.[11] Additionally, participants demonstrate above-average financial literacy skills, answering 73% of the financial literacy questions correctly.[12] Furthermore, 40% were familiar with the concept of ESG before this study, compared to 25% in the German population (IFNP, 2021). On average, participants believe sustainable stocks have about the same future performance but lower risk compared to conventional stocks. Lastly, participants exhibit average pro-social attitudes, donating an average of 44% of their endowment in the standard dictator game and 21% in the "money-burning" dictator game.[13]

[10] We drop two participants from the original sample because we suspect them of having participated twice. All results hold if we include these two.

[11] Stock market participation in Germany is relatively low with approximately 18% of the population aged 14 and above. Furthermore, the role of bonds in this sample is minimal, with only 34 participants reporting holding some bonds, and of those, only 6 exclusively investing in bonds.

[12] Compared to a representative Dutch sample (van Rooij *et al.*, 2011) where participants answer on average 54% and participants of a study with a German fintech company (Hett *et al.*, 2022) where participants answer on average 63% of largely similar questions correctly.

[13] Umer *et al.* (2022) conduct a meta-analysis and find that, on average, participants give 40-60% of their endowment in standard dictator games with charities. In comparison, participants in Crumpler and Grossman

Table 1: Summary statistics of the main variables for the student sample

	Mean	SD	Min	Max	N
ESG Score (allocation 1)	22.86	6.78	5.66	36.87	643
ESG Score (allocation 2)	21.94	5.91	2.34	34.76	643
ESG Score (allocation 3)	20.95	8.92	-6.07	33.67	643
% SRI (allocation 4)	57.78	30.71	0.00	100.00	643
non altruist	0.16	0.36	0.00	1.00	643
warm glow altruist	0.49	0.50	0.00	1.00	643
impact altruist	0.36	0.48	0.00	1.00	643
WTP	3,375.14	2,448.06	0.00	10,000.00	643
age	24.23	4.07	16.00	62.00	643
student	0.83	0.37	0.00	1.00	643
lower SRI returns	3.80	2.01	0.00	7.00	643
lower SRI risk	4.77	1.76	0.00	7.00	643
risk aversion	0.46	0.14	0.00	0.94	643
income	2.02	1.17	1.00	5.00	643
female	0.63	0.48	0.00	1.00	643
financial literacy	7.34	1.97	0.00	10.00	643
inv. experience	0.46	0.50	0.00	1.00	643
sample Mainz	0.50	0.50	0.00	1.00	643
ESG knowledge	0.40	0.49	0.00	1.00	643
warm glow first	0.50	0.50	0.00	1.00	643
receive ESG info	0.32	0.47	0.00	1.00	643
overperformance	0.49	0.50	0.00	1.00	643
positive incentive	0.50	0.50	0.00	1.00	643
moral bad	0.58	0.50	0.00	1.00	332

Notes: Table 19 presents summary statistics. Definitions of these variables can be found in Table 7 (appendix). We impute *risk aversion* with the median risk aversion value for those 25 individuals for whom we could not calculate risk aversion due to multiple switching points in the risk elicitation task. Additionally, since only one treatment group could select a story (*moral bad*), the number of observations in this category is reduced.

2.3 Results

Our primary objective is to shed light on the questions of what drives sustainable investments and whether frictions prevent retail investors from (efficient) sustainable investments. Unless stated otherwise, all analyses were pre-registered before data collection at the American Economic Association's registry for randomized controlled trials under AEARCTR-0008178. Please remember that, given the definition of ESG Scores in our setting, higher ESG Scores indicate greater sustainability.

2.3.1 Sustainability Preferences and the Willingness to Pay for ESG Score Information

The Value of Sustainability for Investors: Our analysis starts with providing three pieces of evidence showing that investors value sustainability: First, in portfolio allocation 1, all par-

(2008) give on average 20% of their endowment in money-burning dictator games with a charity as the recipient.

Table 2: Sustainable investment choices are correlated with altruistic preferences.

	(1) ESG Score	(2) ESG Score	(3) ESG Score	(4) WTP	(5) WTP
receive ESG info	5.161***				
	(0.504)				
WTP	0.258				
	(0.268)				
lower SRI returns	-0.381*	-0.551*	-0.543*	73.32	73.89
	(0.219)	(0.288)	(0.288)	(98.45)	(98.57)
lower SRI risk	0.221	0.141	0.128	-8.851	-9.817
	(0.234)	(0.329)	(0.331)	(97.75)	(98.28)
risk aversion	0.398*	0.687**	0.680**	-12.83	-13.36
	(0.212)	(0.269)	(0.270)	(107.7)	(107.7)
income	0.132	-0.130	-0.130	-55.05	-55.03
	(0.230)	(0.288)	(0.288)	(96.64)	(96.70)
female	0.748	1.115*	1.115*	213.4	213.3
	(0.517)	(0.629)	(0.630)	(208.2)	(208.4)
financial literacy	0.184	0.343	0.312	-65.86	-68.16
	(0.238)	(0.306)	(0.308)	(111.7)	(112.8)
inv. experience	-0.210	0.00344	-0.0248	-778.4***	-780.5***
	(0.505)	(0.602)	(0.605)	(204.4)	(205.5)
sample Mainz	-0.199	0.272	0.288	270.4	271.6
	(0.447)	(0.580)	(0.579)	(205.9)	(205.6)
ESG knowledge	-0.212	-0.386	-0.398	-402.0*	-402.9*
	(0.530)	(0.615)	(0.615)	(208.4)	(208.8)
warm glow first		0.790	0.913	28.54	37.74
		(0.544)	(0.566)	(185.3)	(192.6)
altruist		2.178***		1462.1***	
		(0.820)		(263.2)	
impact altruist			2.468***		1483.7***
			(0.901)		(291.9)
warm glow altruist			1.943**		1444.5***
			(0.850)		(278.4)
Constant	20.11***	19.95***	19.91***	2377.3***	2374.2***
	(0.608)	(0.939)	(0.939)	(308.3)	(307.7)
N	643	643	643	643	643
adj. R squared	0.195	0.0289	0.0285	0.119	0.117

Notes: Table 2 presents different OLS specifications showing (i) that receiving ESG Score information is significantly associated with sustainable investments and (ii) altruistic preferences are significantly related to socially responsible investments and the WTP for ESG Score information. The dependent variables are the mean ESG Score of subjects' portfolio allocation 2 (Specification (1)), mean ESG Score of subjects' portfolio allocation 1 (Specifications (2) - (3)), and the WTP for ESG Score information (Specifications (4) - (5)). The most important explanatory variables are described as follows: *receive ESG Score*: An indicator variable that takes the value one if the individual receives the ESG Score information in portfolio allocation 2; *lower SRI returns*: The investor's response to the statement "I expect shares with a high ESG Score to underperform conventional shares."; *lower SRI risk*: The investor's response to the statement "I expect shares with a high ESG Score to have lower price fluctuations than conventional shares."; *altruist*: An indicator variable that is one if the participant donates something in at least one of the dictator games; *impact altruist*: An indicator variable that is one if the individual donates exclusively in the standard dictator game; *warm glow altruist*: An indicator variable that is one if the participant donates in both dictator games. See Table 7 (appendix) for detailed descriptions of all variables. Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

ticipants receive ESG Scores along with various other pieces of information. The average *ESG Score* from this baseline investment decision is 22.86, significantly exceeding the expected score of 18.33 for a randomly selected portfolio (t-test; $p < 0.01$). This result is robust to other definitions of sustainability (*# sustainable*; *# sustainable d.*; $p < 0.01$). Second, in Stage 2 of our experiment, some participants receive the ESG Score information and can use it to build their portfolio, while others do not. Specification (1) in Table 2 demonstrates that receiving the ESG Score information significantly increases the mean ESG Score of the portfolios. This specification presents the results of an OLS regression in which the dependent variable is the mean ESG Score of portfolio allocation 2. The explanatory variable is an indicator variable that takes the value of one if an investor received the ESG Score information and zero otherwise, while we control for the *WTP* for ESG Scores, risk and return expectations, and several demographic variables.[14] Having the opportunity to trade based on the ESG Score information is correlated with an increase in the ESG Score of the selected portfolio of on average 5.19 points ($p < 0.01$). This translates into an increase of 24% compared to the mean ESG Score of all individuals in Stage 2 of the experiment. This result is robust to changing the sustainability measure as the dependent variable and for different specifications of the control variables (Table 8, appendix). Lastly, 53% of the participants state that they currently own or plan to acquire sustainable investments within the next three years. These results are in line with the literature that establishes investors' sustainability preferences (see e.g., Barreda-Tarrazona *et al.* 2011; Hartzmark and Sussman 2019; Pástor *et al.* 2020; Bauer *et al.* 2021). Taken together, the evidence raises the question: Do investors have a WTP for sustainability information?

Investors WTP for ESG Score Information: We measure the direct WTP for ESG Score information that enables participants to invest sustainably. Our analysis suggests that participants in our sample are willing to pay on average 3,375 points, or 6.75% of their endowment, for sustainability information, which is significantly more than zero (t-test; $p < 0.01$). This result is in line with the literature stating that investors are willing to forgo returns and pay higher fees to invest sustainably (Gutsche and Ziegler, 2019; Heeb *et al.*, 2023; Baker *et al.*, 2022; Laudi *et al.*, 2023; Engler *et al.*, 2023).

Result 1: Investors demonstrate a preference for sustainability and are willing to pay for ESG Score information that enables them to make sustainable investment decisions.

Having established participants' sustainability preferences and their WTP for information that enables them to invest sustainably, the question is whether these preferences are effectively translated into actual sustainable investments, or whether there are real-world frictions that prevent an effective transmission into actual investments. To answer that broader question we start by analyzing what drives SRI. The literature identifies financial (Døskeland and Pedersen,

[14]This part of the analysis is exploratory and adds evidence that investors in our sample value sustainability.

2016; Dong *et al.*, 2022; Giglio *et al.*, 2023) as well as non-pecuniary motivations (Riedl and Smeets, 2017; Bauer *et al.*, 2021; Heeb *et al.*, 2023; Gutsche *et al.*, 2023) as potential factors influencing the decision to invest sustainably. In order to identify the key drivers of sustainable investments, we begin by analyzing the impact of altruistic preferences on sustainable investments, while controlling for risk and return expectations. Subsequently, we delve deeper into the effects of financial motivations.

2.3.2 Altruistic Preferences, SRI and the WTP for ESG Score Information

Given that regulators want to use the financial markets to finance the green transition of the economy, understanding the distinction between warm glow and impact altruistic preferences is essential. Key studies highlight altruistic preferences as crucial for sustainable investment decisions (Riedl and Smeets, 2017; Brodback *et al.*, 2019; Bauer *et al.*, 2021; Gutsche *et al.*, 2023; Heeb *et al.*, 2023). What remains unclear is whether sustainable investments are mainly driven by warm glow or impact altruistic preferences. If primarily impact altruists invest sustainably, investment product providers should highlight the tangible impacts of their products, such as contributions to sustainable development goals, emission reductions, or improved labor conditions (Gutsche *et al.*, 2023). Hence, there is no need for regulators to act, as sustainable investors focus on the impact of the investments. Therefore, greenwashing attempts will be identified by the investors and screened out of the market.[15] On the other hand, if predominantly warm glow altruists buy sustainable investments, the focus of financial institutions should be on products that are labeled and perceived as sustainable and therefore provide emotional satisfaction, even if the actual impact is minimal (Gutsche *et al.*, 2023). Hence, regulators should prevent greenwashing attempts by providing clear labels and guidelines that explicitly define what sustainable investments should entail.[16]

Being Altruistic and Pursuing a Sustainable Investment Strategy: Specifications (2) - (3) in Table 2 present OLS regression results with the *ESG Score* of portfolio allocation 1 as the dependent variable, indicating a positive correlation between altruistic preferences and sustainable investments, controlling for risk and return expectations and demographics. Column (2) displays a statistically significant positive correlation between being an *altruist* and the *ESG Score* in the baseline decision of the stock market experiment ($p < 0.01$). Altruistic individuals hold portfolios with 2.18 higher *ESG Scores* compared to non-altruists, which is 9.5% higher than the mean ESG Score of all individuals. Column (3) categorizes all *altruists* into *impact altruists* and *warm glow altruists*. The results show that both altruistic types are correlated with sustainable investment behavior. Being an *impact altruist* is correlated with 2.47 higher

[15] The withdrawal of products may result from concerns regarding the issuer's reputation or a decline in demand, which ultimately leads to inadequate profits.

[16] The removal of ambiguity eliminates the rationale that warm glow altruists employ to justify questionable investments as sustainable. Consequently, they may be more inclined to invest in genuinely sustainable opportunities.

ESG Scores ($p < 0.01$), while being a *warm glow altruist* is correlated with 1.94 higher *ESG Scores* ($p = 0.02$) compared to non-altruists, ceteris paribus. This translates into increases of 8.5% to 9% compared to the average ESG Score in the baseline portfolio allocation. Wald tests show no significant differences between being an *impact altruist* or a *warm glow altruist* ($p = 0.38$). Specifications (4) and (5) in Table 2 confirm the prior results and demonstrate a sizable correlation between the *WTP* for ESG Score information and being an *impact altruist* (*warm glow altruist*), controlling for demographics and risk and return expectations ($p < 0.01$).

Robustness analyses demonstrate that all results of Specifications (2) - (5) hold for alternative specifications of altruistic preferences (Table 9, Panel A, appendix) and are qualitatively the same for alternative specifications of sustainability (Table 9, Panel B, appendix).[17] We confirm that warm glow motivations play a role in sustainable investment decisions, but overall we find no significant differences between warm glow and impact altruists. Concluding, as a significant portion of investors seems to be motivated by warm glow altruism, regulators must provide clear guidelines for the definition of sustainable (impact) investments to prevent greenwashing attempts that address warm glow altruists.

Result 2: Both, warm glow altruists and impact altruists are more likely to invest sustainably than non-altruists. However, we find no differences in the degree of sustainable investments between warm glow altruists and impact altruists.

2.3.3 Perceived Financial Incentives and SRI

After showing the impact of non-pecuniary motives on sustainable investments, we next address the question of which role financial motives play in sustainable investment decisions. In the first step of our analysis, we take an exploratory approach and follow Riedl and Smeets (2017) to analyze the correlation between return (risk) expectations of SRI and investments in sustainable stocks. Specifications (2) and (3) in Table 2 report a statistically significant negative correlation of the variable *lower SRI returns* with sustainable investments ($p = 0.06$), while the coefficient of *lower SRI risk* is not significant ($p = 0.7$). A one standard deviation reduction in expected returns of sustainable stocks compared to conventional ones decreases the *ESG Score* by 2.4% compared to the average *ESG Score* in the baseline portfolio allocation. This provides an initial indication that financial incentives might influence investment behavior and raises the question: Do investors react to changes in their risk and return expectations of SRI?

According to standard economic theory, investors react to financial incentives (Markowitz, 1952); however, if social factors are the primary motivation for investors to engage in sustainable investments, financial incentives might not only fail to enhance sustainable investment levels but

[17] We have pre-specified the altruism specifications in Table 9, Panel A (appendix) in our pre-analysis plan. However, because we later use the different altruistic types for exploratory heterogeneity analysis, we report them as our main specification.

could even decrease them (Frey and Oberholzer-Gee, 1997; Mellström and Johannesson, 2008; Gneezy *et al.*, 2011; Brodback *et al.*, 2019). This issue becomes particularly significant considering the observed correlation between being altruistic and socially responsible investments. Døskeland and Pedersen (2016) utilize the Levitt and List (2007) framework, which posits that utility derived from financial and social benefits is additively separable, in the context of sustainable investments. Expanding this model to include the potential negative impacts of financial incentives on the social utility associated with sustainable investments highlights four possible responses to performance information: When investors are presented with information showing the superior performance of sustainable stocks, they may (i) increase their sustainable investments as the financial incentives are consistent with their social objectives, thereby deriving utility from both the financial and social aspects of their investments. Alternatively, (ii) higher financial returns could diminish the intrinsic motivation for pro-social behaviors, such as investing sustainably, potentially resulting in unchanged or decreased sustainable investments as the social benefits are negated by the financial gains. This is consistent with the findings of Mellström and Johannesson (2008) in the context of blood donation and Brodback *et al.* (2019) who indicate that altruists' motivation to invest sustainably decreases if they expect higher returns from these investments. Conversely, when exposed to information indicating the underperformance of sustainable stocks, (iii) investors may reduce their sustainable investments if the financial disadvantages significantly outweigh the social benefits. On the other hand, (iv) such information might have no effect or even increase sustainable investment levels if the perceived social benefits surpass the financial drawbacks, as the social value of these investments might rise with the financial disutility.

We exogenously vary the risk-adjusted performance expectations of high-sustainability stocks at two points in the experiment. First, in Stage 3 of the experiment, we introduce an information treatment that exogenously varies the information on the risk-adjusted performance of high-sustainability firms. This is followed by a participation in the stock market game, which allows us to analyze the within-subject change in investment behavior. Second, in Stage 4 of the experiment, we introduce an incentive treatment that exogenously varies the direct financial incentives to invest in a high-sustainability portfolio. Participants subsequently allocate their endowment between a sustainable and a non-sustainable portfolio. Table 10, Panels A and B (appendix), demonstrates that the randomization between the treatments overall worked. To address concerns regarding differences in sustainability preferences before the information treatment in Stage 3 (Panel A), we employ an individual fixed effects regression. Most importantly, we observe no differences in sustainability preferences for the incentive treatment in Stage 4 (Panel B).

Reaction to the Information Treatment: We start by analyzing the effect of the information treatment. To address minor pre-treatment differences in baseline sustainability demand

and show the impact of within-subject reactions to overperformance and underperformance information, we apply an individual fixed effects regression model as the preferred specification:

$$ESG_{it} = \alpha + \beta_1 \cdot \text{overperformance}_{it} + \beta_2 \cdot \text{underperformance}_{it} + \gamma_i + \epsilon_{it} \quad (1)$$

where ESG_{it} represents the mean *ESG Score* for individual i at time t, *overperformance* and *underperformance* are treatment assignment indicator variables, γ_i represents individual fixed effects, and ϵ_{it} is the error term.

Table 3: Participants' investment behavior reacts to information regarding the overperformance and the underperformance of sustainable stocks.

	(1) ESG Score	(2) ESG Score	(3) ESG Score	(4) ESG Score
overperformance	7.035***	7.065***	6.277***	7.531***
	(0.426)	(1.235)	(0.724)	(0.577)
underperformance	-2.973***	-0.922	-2.456***	-4.106***
	(0.482)	(1.163)	(0.780)	(0.712)
Constant	20.95***	18.78***	21.53***	21.21***
	(0.161)	(0.423)	(0.269)	(0.228)
N	1286	200	462	624
fixed effects	yes	yes	yes	yes
sample split	-	non altruists	impact altruists	warm glow altruists

Notes: Table 3 uses individual fixed effects regressions showing that investors' investment behavior significantly reacts to receiving risk-adjusted overperformance and underperformance information. The dependent variable, *ESG Score*, is the mean ESG score of individual i at time $t \in \{1, 3\}$ (baseline and post-treatment portfolio allocations). *Overperformance* is an indicator variable set to one if individual i received the overperformance information at any time t, and zero otherwise. Similarly, *Underperformance* indicates whether individual i received underperformance information at any time t. Specifications (2) - (4) are exploratory; Specification (2) includes only altruistic investors, while Specification (3) is restricted to impact altruists and Specification (4) comprises exclusively warm glow altruists. Significance levels are indicated by *, **, and *** for the 10, 5, and 1 percent levels, respectively.

Table 3 presents the regression outcomes. Column (1) indicates that exposure to ESG overperformance (underperformance) information significantly increases (decreases) *ESG Scores* relative to the baseline portfolio allocation of the same individual ($p < 0.01$), with changes of 33.6% (14.2%) respectively. In an exploratory analysis, we evaluate if all types of altruists react to performance information. Specifications (2) - (4) demonstrate a significant increase in ESG holdings after receiving overperformance information for non-altruists, impact altruists, and warm glow altruists ($p < 0.01$). Contrarily, we observe a significant decrease in *ESG Scores* after receiving underperformance information for impact altruists (Specification (3); $p < 0.01$) and warm glow altruists (Specification (4); $p < 0.01$), while there is no significant reaction of non-altruists (Specification (2); $p = 0.43$). The most likely reason is that non-altruists already hold significantly less sustainable portfolios in the baseline portfolio allocation compared to altruists (see e.g., Table 2). Taken together, altruistic individuals strongly react to overperformance and underperformance information, while non-altruists only adjust their investment behavior after receiving overperformance information. We therefore find no evidence for crowding out of perceived financial incentives due to decreased social utility.

Given the different underlying motivations to invest sustainably for impact altruists and warm glow altruists, we exploratory analyze if warm glow altruists react stronger to performance information than impact altruists (and non-altruists). Table 12 (appendix) displays an OLS regression using the difference between pre- and post-treatment *ESG Scores* as the dependent variable and the interaction of being in the overperformance treatment and being a warm glow altruist as the variable of interest. The interaction effect shows that warm glow altruists react significantly stronger to the information treatment than impact altruists (Specification (1); $p = 0.04$) and all non-warm glow altruists (Specification (2); $p = 0.02$). This finding is robust to different measures of sustainability (Specifications (3) - (6); all $p < 0.06$). One interpretation of this finding is that warm glow altruists want to feel good about their investment but do not care about the true impact, therefore the financial gain of not investing sustainably in the underperformance treatment might outweigh the decreasing social utility of not investing sustainably. On the other hand, they react strongly to the benefits of the overperformance treatment as they get high financial returns while still feeling good about investing sustainably.

Reaction to the Incentive Treatment: For the last part of the evidence on the role of (perceived) financial incentives on SRI, we exogenously vary the direct financial incentive to invest in a high-sustainability portfolio. Figure 3 shows the share of SRI for every investor, broken down by each treatment (only no study group to analyze exclusively the incentive effect). Two things are worth noting. First, although participants in the negative investment incentive treatment get deducted ten percent of their endowment which they invest in the high-sustainability portfolio, a substantial number of investors still have a large preference for SRI. Second, providing a positive financial incentive substantially increases SRI. Specification (1) in Table 13 (appendix) demonstrates that receiving a positive financial incentive increases SRI by 55% compared to those who received a negative incentive ($p < 0.01$). This result also remains stable once controls are added (Specification (2)). It is worth noting that the magnitude of the incentive and the information treatment effect is about the same ($p = 0.6310$) for participants in the no-study group.[18] Table 14 (appendix) confirms the previous result that investors react to incentives and adds that investors react significantly differently to various levels of positive and negative incentives ($p < 0.01$).

[18]We repeat this analysis for the study group and show that while the incentive effect remains about the same, the risk-adjusted information effect decreases to about half ($p < 0.01$). This is most likely due to the interaction between the overperformance information and the study on the moral consequences of sustainable investments. Moreover, incentives also shift which study is endogenously chosen and how incentives change how the information is processed (see subsection 2.3.5).

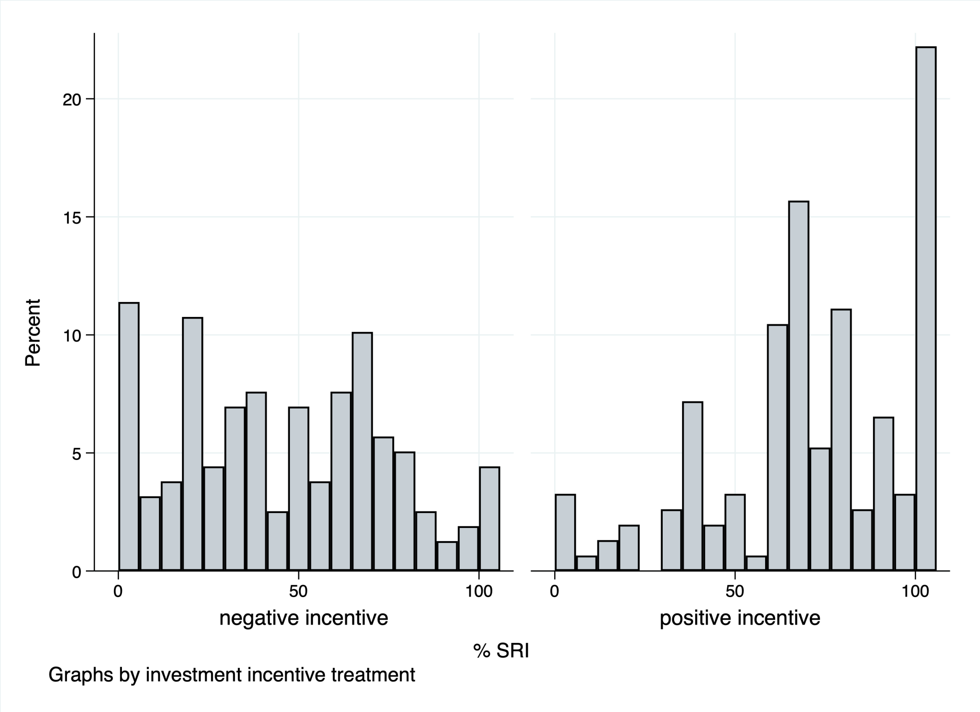


Figure 3: Direct financial incentives matter for sustainable investment decisions.
Notes: This graph displays the share of investments in the sustainable portfolio for each investor in the positive (negative) incentive treatment. Investors in the positive incentive treatment receive a subsidy of 20% on their investment in the sustainable portfolio, while investors in the negative incentive treatment are deducted 20% of their investment in the sustainable portfolio. We exclude all participants with the possibility to pick a story on the moral consequences of SRI to focus on the pure incentive effect.

Taken together, we find no evidence supporting the crowding out hypothesis for SRI. On the contrary, we establish a strong positive causal effect of (perceived) monetary incentives on SRI. Shifting participants' beliefs about the risk-adjusted financial consequences of SRI has a significant and economically meaningful impact on their investment decisions. This contradicts the literature, which states that social preferences drive SRI while financial factors only play a minor role (e.g., Riedl and Smeets 2017; Bauer *et al.* 2021). One reason for this might be that existing research usually relies on survey questions to ask about risk and return expectations and measures their correlation with SRI. Dong et al. (2022) show that investors in self-reported non-incentivized survey questions systematically understate their expectations on the performance of sustainable assets to create a positive self-image via social signaling. In Table 2, we replicate the most common approach in the literature and show a marginally significant effect of return expectations on SRI holdings. On the contrary, in the information as well as in the incentive treatment, we exogenously shift the beliefs of risk and return expectations of sustainable investments and therefore infer the direct effect of changed beliefs on SRI. To conclude, risk and return expectations of SRI play a crucial role in the investment process. Participants react to both performance information and direct financial incentives in line with standard economic theory.

Result 3: Investors react to (perceived) financial incentives according to standard economic theory. Heterogeneity analysis reveals that warm glow altruists adjust their investment behavior the strongest.

These results raise the question: Can financial markets efficiently implement individuals' sustainability preferences, or are there frictions preventing an effective allocation of capital? Given the high degree of conflicting information in the market for sustainability, this question is highly relevant. As of today, there are no mandatory guidelines that define and label sustainable investments consistently across the world (Filippini *et al.*, 2024). Currently, two major issues exist. First, there is only limited common ground for the reporting of sustainability information for firms, making it challenging to compare the available information between firms.[19] Second, different ESG rating agencies not only use different inputs but also differ in their methodologies, resulting in highly heterogeneous ratings of the same firms between different rating agencies (Dimson *et al.*, 2020; Billio *et al.*, 2021; Avramov *et al.*, 2022; Berg *et al.*, 2022). Moreover, even if we take a certain definition of sustainability as given, the results on the performance of sustainable stocks are mixed (Eccles *et al.*, 2014; Friede *et al.*, 2015; Pástor *et al.*, 2020; Pastor *et al.*, 2021; Avramov *et al.*, 2022; Bolton and Kacperczyk, 2023; Latino, 2023; Bolton and Kacperczyk, 2024). Taken together, various frictions currently exist in the market for sustainable products.

There are four possible outcomes of these frictions, each with different implications for research and policy, depending on (i) whether people selectively acquire information and (ii) whether people use the acquired information accordingly. First, if investors acquire information independently of their beliefs about the risk and return expectations of SRI and do not use the acquired information in any way, there is no need for action. Second, if investors acquire information independently of their beliefs about the risk and return expectations of SRI and subsequently use the acquired information to inform their investment decisions, the ambiguous information induces noise in the investment decisions. Therefore, we argue for transparency and clear rules for ESG rating agencies. Third, if investors selectively acquire information but do not use this information for further trading, they rationalize their planned (non-) SRI decision by selectively acquiring information. Hence, resolving the frictions won't affect individuals' SRI decisions because these people would have made the same decision anyway; therefore, no further actions are needed. Fourth, if investors selectively acquire information and then use this information for their subsequent trades, resolving the information frictions and providing clear data and labels enables an effective transmission of individuals' sustainability preferences.

[19]The Corporate Sustainability Reporting Directive (CSRD) addresses this issue for large European companies as well as non-European companies that generate over EUR 150 million in the EU market. Starting from 2025, companies must publish regular reports on their environmental and social impact activities (starting with the fiscal year 2024) (Commission, 2022b, 2024). Furthermore, the IFRS developed the International Sustainability Standards, which apply from the fiscal year 2024 onwards and are intended to provide guidance for a global baseline in sustainability disclosure (IFRS, 2024). The standard is not mandatory, and although the IFRS has been quite successful in setting standards in the past, the future uptake of the sustainability standards remains unclear.

This is especially important because information frictions not only induce noise into individuals' decision-making but also introduce a behavioral mechanism that prevents people from making the decisions they would have made without the frictions. In the next step, we test these four different scenarios, starting with the question of whether individuals acquire information opportunistically.

2.3.4 Incentives and Information Acquisition

For this part of the analysis, we only include the 50% of our participants who can endogenously choose between two research articles: one explaining the moral advantages of SRI (henceforth referred to as the "moral good article") and the other stating the moral disadvantages of SRI (hereinafter referred to as the "moral bad article"). By making the choice of the research article endogenous, we operationalize the real-world situation where information on every possible point of view is available (see the discussion in the previous subsection), and people can select which information they want to read and subsequently include in their decision-making process.

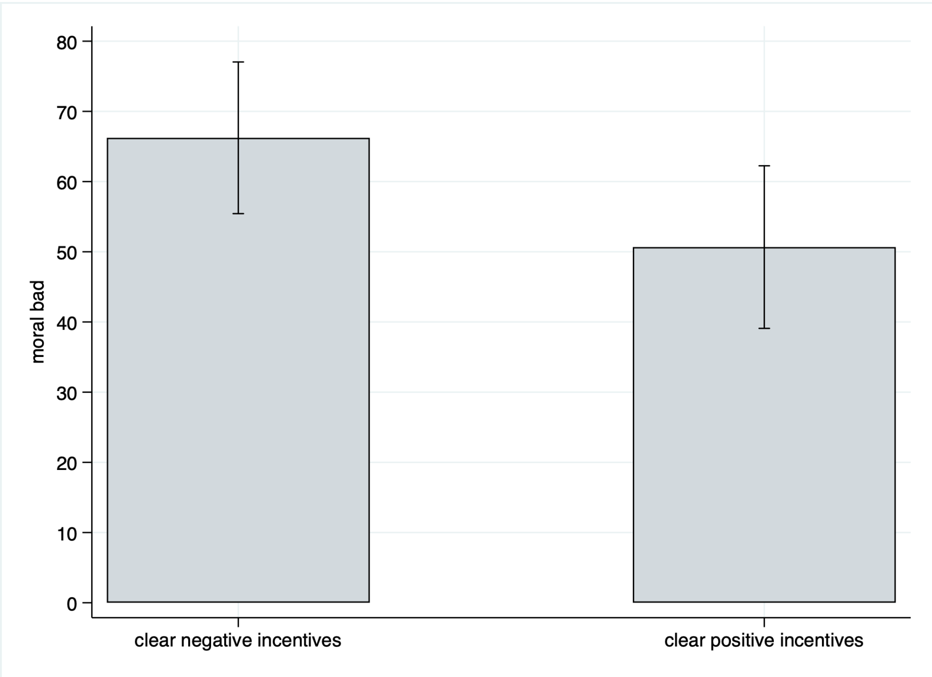


Figure 4: Individuals with clear negative incentives are more likely to select the moral bad article.

Notes: This graph displays the proportion of individuals choosing the moral bad story by (perceived) financial incentive. The result shows that participants with clear positive incentives are less likely to choose the moral bad story than participants with a clear negative incentive. Participants with a clear positive incentive are those investors who are part of both the overperformance treatment group and the positive incentive treatment group. On the other hand, participants with a clear negative incentive are part of the underperformance treatment group and the negative incentive treatment group. The whiskers denote the 95% confidence intervals. See Table 15 for the corresponding regression results.

Figure 4 illustrates the mean probability of selecting the moral bad story based on (perceived) financial incentives.[20] Those individuals with clear negative financial incentives to invest in

[20] Deviating from the pre-analysis plan, in our main specification we only include those individuals who have a

SRI (negative investment incentive and underperformance of SRI research article) select the article outlining why it makes no sense to invest sustainably from a moral standpoint with a probability of 66.30%, whereas individuals with clear positive financial incentives (positive investment incentive and overperformance of SRI research article) are 15.6 percentage points less likely to pick the mentioned research article ($p = 0.052$, Specification (1), Table 4). This suggests that individuals with clear negative incentives to invest in high-sustainability firms are 30.80% more likely to pick the moral bad story than those who exogenously receive clear positive incentives.

Table 4: (Perceived) financial incentives skew information acquisition on the moral consequences of sustainable investments.

	(1) moral bad	(2) moral bad	(3) moral bad	(4) moral bad	(5) moral bad
clear positive incentives	-0.156*	0.0333	-0.0397	-0.313***	-0.0170
	(0.0795)	(0.226)	(0.126)	(0.115)	(0.109)
warm glow altruist					0.0965
					(0.109)
clear positive incentives * warm glow altruist					-0.296*
					(0.158)
Constant	0.662***	0.667***	0.611***	0.719***	0.622***
	(0.0543)	(0.166)	(0.0825)	(0.0807)	(0.0732)
N	152	19	64	69	152
adj. R squared	0.0185	-0.0575	-0.0145	0.0852	0.0307
sample split	-	non altruists	impact altruists	warm glow altruists	-

Notes: Table 4 presents an OLS regression showing that individuals receiving clear positive incentives are less likely to select the article outlining the moral shortcomings of SRI. *Moral bad* refers to a dummy variable that is one if an individual selects and reads the article *"Why investing in sustainable stocks makes no sense from a moral point of view"*. The *clear positive incentives* dummy variable is equal to one if the investor is part of both the overperformance treatment group and the positive incentive treatment group. The indicator variable is zero if the individual is part of the underperformance treatment group and the negative incentive treatment group. *Warm glow altruist* is an indicator variable that is one if the participant donates in both dictator games. Lastly, the *clear positive incentive * warm glow altruist* is the interaction effect of the two previously specified variables. Specifications (1) and (5) include all individuals, while Specifications (2) - (4) focus on non-altruists, impact altruists, and warm glow altruists respectively. Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

Exploratory heterogeneity analysis shows that this effect is entirely driven by warm glow altruists. While non-altruists (Specification (2), $p = 0.88$) and impact altruists (Specification (3), $p = 0.75$) do not react to (perceived) financial incentives, warm glow altruists react strongly ($p < 0.01$). Specification (5) demonstrates that warm glow altruists are 48% less likely to select the moral bad research article if they perceive financial incentives to do so, compared to all

clear positive or a clear negative incentive. The reason is that we do not know how unclear (perceived) financial incentives (positive financial incentive and negative risk-adjusted performance information or vice versa) influence individuals' aggregated beliefs about the financial consequences of SRI. The effects of the pre-registered version are in the expected direction but not statistically significant (Specification (3) in Table 15, appendix; $p = 0.25$). This effect is driven by warm glow altruists (Specification (2)).

other individuals who receive negative financial incentives to invest sustainably ($p = 0.06$). To conclude, we find evidence for the hypothesis that motivated beliefs influence the information acquisition process. Warm glow altruists use the moral 'wiggle room' to selectively acquire the information which is most in line with their (perceived) financial incentives. This is in line with the idea that they want to feel good about their investment decision. Selecting information that tells them that sustainable investments are morally bad anyway can help them justify their subsequent investment decision.

Result 4: Participants opportunistically acquire the information on the sustainability of investments that are most in line with their (perceived) financial incentives. This effect is driven by warm glow altruists.

This result raises the question of the effect of skewed information acquisition on SRI. Do people just pick the information most in line with their financial incentives and then follow through with the investment decision they wanted to make anyway, or do participants actively use the acquired information? If the latter is true, information frictions prevent people from making sustainable investment decisions, and policy actions are required.

2.3.5 Reaction to Information on Moral Consequences

To answer this question, we begin by exploratory analyzing the correlation between choosing the moral bad article and SRI. Individuals who select the moral bad article invest 12.69 percentage points less in the sustainable portfolio than their counterparts ($p < 0.01$), which translates into a decrease in sustainable investments of 19.5%. There could be two different reasons for this result. First, participants might actively select the information that aligns with the decision they would have made anyway to justify their behavior. If this is the case, we should find no effect of actually reading the research articles on the final investment decision. Second, participants might use the information to justify their behavior, and by reading this research article, they further convince themselves (not) to invest sustainably. To distinguish between these two scenarios, we use a nearest-neighbor matching approach including all participants in our analysis.[21]

For every single participant who reads one of the two articles, we search for a perfect match among the participants who did not have the chance to read an article. We use exact matches for the information and incentive treatment as well as for the altruistic type. Additionally, we look for the nearest neighbors in expected ESG return, gender, risk attitude, and financial literacy.[22]

[21] Nearest neighbor matching was not pre-registered; however, it offers two main advantages: First, it allows us to quantify the direct effect of reading the provided stories. Second, due to the interaction between the incentive and the information treatment, we cannot use the original approach. The results of the pre-registered regression can be found in Table 16 (appendix).

[22] We select the altruistic type, *overperformance*, and *positive incentive* as exact matches as these variables have shown a strong impact on SRI and we use altruistic types for heterogeneity analysis. Furthermore, we choose the variables that have been shown to be predictors of SRI in Table 2 (*lower ESG return, risk aversion, female*) as variables for the nearest neighbor matching. To get to our final specification, we enrich the matching with sustainability preferences before reading the story (# sustainable (allocation 2), # sustainable (allocation 3)) and *financial literacy* as the latter shows significant differences between control and treatment groups.

Table 5: Matching was successful (only individuals with clear incentives).

Panel A: Characteristics moral bad story and matched control partners

	moral bad story	matched control	Diff.	Std. Error	Obs.
positive incentive	0.43	0.43	0.00	0.07	178
overperformance	0.43	0.43	0.00	0.07	178
# sustainable (allocation 1)	2.78	2.75	-0.02	0.17	178
# sustainable (allocation 2)	2.22	2.24	0.01	0.18	178
# sustainable (allocation 3)	3.07	3.27	0.20	0.22	178
non altruist	0.15	0.15	0.00	0.05	178
impact altruist	0.43	0.43	0.00	0.07	178
warm glow altruist	0.26	0.26	0.00	0.07	178
partly warm glow altruist	0.17	0.17	0.00	0.06	178
warm glow first	0.52	0.49	-0.02	0.08	178
lower SRI returns	3.79	4.09	0.30	0.32	178
lower SRI risk	4.74	5.11	0.37	0.27	178
risk aversion	0.46	0.47	0.00	0.02	178
income	1.99	2.01	0.02	0.17	178
female	0.56	0.56	0.00	0.07	178
financial literacy	7.80	8.09	0.29	0.24	178
inv. experience	0.48	0.53	0.04	0.08	178
sample Mainz	0.46	0.55	0.09	0.08	178
ESG knowledge	0.46	0.42	-0.04	0.07	178
WTP	3032.03	2796.39	-235.64	308.06	178

Panel B: Characteristics moral good story and matched control partners

	moral good story	matched control	Diff.	Std. Error	Obs.
positive incentive	0.59	0.59	0.00	0.09	126
overperformance	0.59	0.59	0.00	0.09	126
# sustainable (allocation 1)	2.81	3.00	0.19	0.20	126
# sustainable (allocation 2)	2.25	2.11	-0.14	0.22	126
# sustainable (allocation 3)	3.30	3.62	0.32	0.27	126
non altruist	0.10	0.10	0.00	0.05	126
impact altruist	0.41	0.41	0.00	0.09	126
warm glow altruist	0.22	0.22	0.00	0.07	126
partly warm glow altruist	0.27	0.27	0.00	0.08	126
warm glow first	0.44	0.52	0.08	0.09	126
lower SRI returns	4.03	3.90	-0.13	0.30	126
lower SRI risk	4.67	4.94	0.27	0.26	126
risk aversion	0.48	0.46	-0.02	0.02	126
income	2.05	1.94	-0.11	0.21	126
female	0.68	0.68	0.00	0.08	126
financial literacy	6.68	7.38	0.70**	0.33	126
inv. experience	0.38	0.49	0.11	0.09	126
sample Mainz	0.46	0.52	0.06	0.09	126
ESG knowledge	0.30	0.43	0.13	0.09	126
WTP	3579.70	3331.33	-248.37	372.81	126

As our main specification, we only use those individuals with clear incentives because they are biased by selective information acquisition. Table 5 demonstrates that the matching for those reading the moral good story (Panel A) and reading the moral bad story (Panel B) worked. We use t-tests to test for differences in means and identify only minor differences for *financial literacy* for those reading the moral good story, which is not a significant predictor of sustainable investments in our sample (see Table 2). All other variables are not significantly different at conventional levels of significance. The matching quality also holds for the full sample of individuals reading a study on the moral consequences of sustainable investments (Table 17, appendix). To analyze the effect of reading the research articles, we begin by calculating the average treatment effect of the treated (ATET) for each participant in the study group. The ATET reflects the difference in the share of sustainable investments in portfolio allocation 4 of an individual reading the moral bad (good) research article compared to the matched control not reading any article at all. In a second step, we analyze the difference in the ATET between participants who read the moral good research article and those who read the moral bad research article. To achieve this, we employ an OLS regression with the ATET as the dependent variable and an indicator variable that is one if the participant reads the moral bad research article and zero otherwise (Table 6).

Table 6: Reading articles on the moral consequences of sustainable investments influences investment decisions.

	(1) ATET	(2) ATET	(3) ATET	(4) ATET
reading moral bad	-13.58***	20.37	-18.17**	-18.40***
	(4.925)	(19.44)	(7.545)	(6.365)
Constant	2.421	-12.83	2.827	5.032
	(3.769)	(16.08)	(5.814)	(4.723)
N	152	19	64	69
sample split	-	non altruist	impact altruist	warm glow altruist
matching	nnm	nnm	nnm	nnm

Notes: Table 6 displays an OLS regression showing that reading the moral bad article significantly reduces the share of sustainable investments compared to reading the moral good article. The dependent variable *ATET* is the average treatment effect on the treated of reading the moral bad (good) article. We calculate the ATET for each participant in the study group by subtracting the SRI share of the matched control (nearest neighbor matching) from the SRI share of the respective participant in portfolio allocation 4. The explanatory variable *reading moral bad* is an indicator variable set to one if the participant reads the moral bad research article and zero if the participant reads the moral good article. Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

Our results offer evidence that reading the research article discussing the moral downsides of sustainable investing decreases the SRI holdings by 13.58 percentage points compared to reading the research article focusing on the moral upsides of sustainable investments (Specification (1); $p < 0.01$). This effect is sizable given the average share of sustainable investments of 57%. While non-altruists show no significant reaction to reading a research article (Specification (2);

$p = 0.31$), this result is driven by impact altruists (Specification (3); $p = 0.02$) and warm glow altruists (Specification (4); $p < 0.01$). This aligns with the interpretation that non-altruists do not care about the impact of their investment, while impact altruists significantly decrease their sustainable investments if they learn that these investments have no positive impact on society or the environment. Furthermore, warm glow altruists significantly decrease sustainable investments when learning about the negative moral consequences of sustainable investments, as this affects the emotional reward of doing a good deed. Table 18 (appendix) demonstrates that our results are robust to including all individuals in the sample.

In unreported specifications, we explore other methods such as propensity score matching and inverse probability weighting. The quality of the match for the nearest neighbor specification was most convincing, largely due to the strong influence of the information and the incentive treatment on investments in SRI. Among these specifications, only nearest-neighbor matching allows for perfect matches. Consequently, we can conclude that participants not only respond to (perceived) financial incentives but also react to moral information. This reaction is driven by both impact altruists and warm glow altruists.

Result 5: People react to information on the moral consequences of sustainable investments. This is driven by warm glow altruists and impact altruists.

Taken together, Results 4 and 5 demonstrate a behavioral mechanism that influences investors' portfolio choices. Our data shows that individuals select the information most in line with their (perceived) financial incentives (Result 4) and subsequently act on the acquired information (Result 5). Specifically, warm glow altruists drive the effect of selective information acquisition. Our results show that the exact same individuals subsequently also react to reading the story on the moral consequences of sustainable investing. Hence, their skewed information acquisition influences subsequent investment behavior. We advocate for resolving information frictions and providing clear data and labeling in the market of sustainable investments to enable the effective transmission of true sustainability preferences. The current regulatory framework allows investors to explicitly avoid costly information. This skewed information acquisition, in turn, persuades investors not to invest sustainably.

Result 6: Opportunistic information acquisition on the moral behavior of sustainable investments changes subsequent investment behavior. This effect is driven by warm glow altruists.

In general, financial markets have all the tools to implement investors' sustainability preferences. However, in the market for sustainable investment products, numerous information frictions prevent the effective transmission of these preferences. For example, investors notice the costs of sustainable investments and the non-transparency of sustainability indicators. This amplifies the issue that investors do not invest sustainably, as they can actively search for information that justifies their non-SRI decision and then use this information to consolidate their decision and invest less sustainably. In the last step, we perform a back-of-the-envelope calculation to quantify the effect of this behavioral mechanism and compare the impact of the switch from the moral good to the moral bad story with the needed financial incentives to achieve the same result. Assuming a linear relationship between incentive and investments in SRI, Table 14 (appendix) demonstrates that a direct financial inventive of 1% leads to a 0.817 percentage point increase in SRI. [23] This number enables us to quantify the monetary effect of our information treatment and the behavioral mechanism explained in the section above. According to Table 13, reading the risk-adjusted overperformance research article increases SRI by 23.73 percentage points compared to reading the risk-adjusted underperformance research article. Therefore, this effect equals a financial incentive to invest sustainably of 19.39 percentage points.[24] Finally, we quantify the financial equivalent of the behavioral mechanism discussed in the previous section. Following Table 6, we conclude that reading the moral bad research article decreases the share of sustainable investments by 13.58 percentage points compared to reading the moral good article. Taken together, the shift from the moral bad to the moral good story has an effect equivalent to a monetary incentive on the investments in sustainable stocks of 12.73 percentage points.

2.4 Discussion and Conclusion

Given the attempts to use the financial markets to stimulate the green transition (see e.g., European Commission (2019a)), it is highly relevant to understand which factors drive sustainable investments. To investigate this question, we conduct an online experiment with different incentive and information treatments and observe investors' reactions to these treatments in an incentivized stock market game. First, we show that participants in our sample have a preference for SRI and are willing to pay for sustainability information. We demonstrate that both moral factors and (perceived) financial incentives drive sustainable investments. Second, our results illustrate that investors opportunistically acquire information on the moral consequences of SRI, which enables a mechanism of strategic self-deception. People not only skew their information acquisition to justify their investment decisions but also use the obtained information on the moral consequences of SRI to further convince themselves of their (non-) SRI decisions. This skewed information acquisition is driven by warm glow altruists and prevents an efficient transmission of existing sustainability preferences. These results are not artifacts of the sample composition in this paper. Braun *et al.* (2024c) show in a similar experiment with wealthy private investors that these investors also exhibit substantial preferences for sustainability (and are willing to pay for sustainability information) which is correlated with altruistic preferences.

[23] We take the most conservative approach here. The switch from no incentive to a -10% incentive is associated with a decrease in SRI of 20.41%.

[24] We restrict the analysis to the no study group to exclude interactions between the overperformance research article and the selected study on the moral behavior of SRI.

Moreover, they document a significant reaction to risk-adjusted performance information of high sustainability assets.

We believe our results may be relevant for academics as well as policymakers and practitioners. Given the political debate on how to increase socially responsible investments, we suggest different avenues that might be successful. Our results show a positive relationship between (perceived) financial incentives and SRI. Hence, tax discounts on returns or subsidies on the costs of sustainable investments could be a way to increase SRI. Additionally, we provide evidence that participants care about the moral implications of their investments. However, a significant fraction of sustainable investors are warm glow altruists susceptible to greenwashing. Therefore, transparent data, consistent ESG ratings, and salient information on the moral consequences of SRI are key to enabling optimal capital allocation and preventing strategic self-deception. The labeling of sustainable products in the Sustainable Finance Disclosure Regulation (Commission, 2019b) is a step in the right direction. However, as shown in Braun *et al.* (2024a), it remains challenging for investors to identify genuine impact investments as they tend to rationalize investments as sustainable if they align with their financial interests. Moreover, the issue of conflicting ESG ratings remains as rating providers use different methodologies to calculate their ratings (Dimson *et al.*, 2020; Billio *et al.*, 2021; Berg *et al.*, 2022), with the consequence that investors potentially pick the rating most in line with their underlying motivations. For the aforementioned reasons, we advocate for the implementation of regulations that reduce the ambiguity surrounding the definition of sustainable investments. We believe that providing investors with a set of sustainability definitions (as evidenced by the amendment to the MiFID II regulation (see ESMA (2023)) is not a viable solution.

Our results also demonstrate a behavioral mechanism that may prevent investors from investing sustainably due to negative beliefs about the financial consequences of SRI. Our findings suggest that investors selectively acquire information that aligns with their beliefs on the financial returns of SRI. This indicates that a targeted financial incentive can work in two ways. First, the incentive may have a direct effect on the willingness to invest sustainably. Second, the incentive may have an indirect effect by changing how people acquire information, leading investors not only to react to the incentive but also to use the acquired information (which aligns with the incentive) to convince themselves to invest sustainably. Given the large amount of conflicting information in the area of SRI, changing people's beliefs about the financial consequences of SRI may be a powerful tool to increase SRI.

Future research should analyze the costs and benefits of information campaigns compared to financial incentives in a real-world investment case and shed light on the question of what is most cost-effective to boost SRI. Furthermore, it would be interesting to analyze the behavioral mechanism described in this paper with a set of (professional) real-world investors who are confronted with conflicting ESG ratings. Taken together, utilizing financial markets to unleash consumer demand for SRI might be an effective tool for improving the sustainability of economic activity. However, investors exploit opportunities to rationalize non-SRI investments. Transparent data and information are key to enabling optimal capital allocation and preventing strategic self-deception.

2.5 Appendix

Table 7: Variable definitions

Variable	Description
ESG Score	The average ESG Score of an investor's portfolio choice in allocations 1, 2, or 3.
# sustainable	Number of sustainable stocks in an investor's portfolio for allocations 1, 2, or 3.
# sustainable d.	Dummy variable equal to one if the participant invests in more than two sustainable stocks in an investor's portfolio for allocations 1, 2, or 3.
% SRI	The percentage share of investments in the sustainable portfolio for allocations 4-8.
WTP	The amount of points an investor is willing to spend to obtain ESG Scores in portfolio allocation 2 (range 0-10,000).
moral bad	Dummy variable that is one if the individual selects the article "Why investing in sustainable stocks makes no sense from a moral point of view."
ATET	Average treatment effect on the treated of reading the moral bad (good) article. Calculated by subtracting the SRI share of the nearest neighbor matched control from the SRI share of the respective participant in portfolio allocation 4.
receive ESG Score	Indicator variable that is one if the individual receives the ESG Score information in portfolio allocation 2.
altruist	Indicator variable that is one if the individual donates at least one euro in at least one of the two dictator games.
non-altruist	Indicator variable that is one if the individual donates less than one euro in both of the two dictator games.
impact altruist	Indicator variable that is one if the individual donates at least one euro in the standard dictator game and less than one euro in the "money-burning" dictator game.
warm glow altruist	Indicator variable that is one if the individual donates at least one euro in the money-burning dictator game.
altruism	Amount of the donation to the charity in the standard dictator game (between 0 and 8 euros).
impact altruism	Difference between the donation amounts in the standard and "money-burning" dictator games.
warm glow altruism	Amount of the donation in the "money-burning" dictator game (between 0 and 8 euros).
warm glow first	Dummy variable equal to one if the investor participated in the impure altruism dictator game task before the warm glow dictator game task.
overperformance	Dummy variable for participation in the overperformance treatment.
underperformance	Dummy variable for participation in the underperformance treatment.
positive incentive	Dummy variable for participation in the treatment where investments in the sustainable portfolio are multiplied by 1.1.
clear positive incentives	Dummy variable equal to one if the investor is in the overperformance and positive incentive treatments.
unclear positive incentives	Dummy variable for investors in either the overperformance and negative incentive treatment or the underperformance and positive incentive treatment.

Table 7 (continued):

Variable	Description
reading moral bad	Dummy variable set to one if the participant reads the morally negative research article and zero if the participant reads the morally positive article.
lower SRI returns	The investor's response to the statement "I expect shares with a high ESG score to underperform conventional shares," on a scale from 0 (does not apply at all) to 7 (applies completely).
lower SRI risk	The investor's response to the statement "I expect shares with a high ESG score to have lower price fluctuations than conventional shares," on a scale from 0 to 7.
ESG knowledge	Dummy variable equal to one if the investor is familiar with the concept of ESG before the study.
risk aversion	Indicates the investor's risk attitude as measured in the risk elicitation task, where a higher value indicates more risk aversion (scale 0-0.94).
financial literacy	The number of correctly answered questions in a financial literacy quiz (0-10).
female	Dummy variable equal to one if the investor identifies as female.
age	The investor's self-reported age.
income	The investor's response to "How high is your personal net income per month?" categorized as: (1; 0-700 euro) (2; 701 - 1,000 euro) (3; 1,001 - 1,500 euro) (4; 1,501 - 2,500 euro) (5; more than 2,500 euro)
inv. experience	Dummy variable equal to one if the investor holds bonds, stocks, or funds in the real stock market.
sample Mainz	Dummy variable equal to one if the investor participated in the study via the Johannes Gutenberg University Mainz (MABELLA).

Notes: Table 7 defines all relevant variables.

Table 8: Investors consider ESG Scores in their investment decisions.

	(1) ESG Score	(2) ESG Score	(3) # sustainable	(4) # sustainable	(5) # sustainable d.	(6) # sustainable d.
receive ESG info	5.596***	5.161***	1.241***	1.152***	0.459***	0.427***
	(0.483)	(0.504)	(0.104)	(0.113)	(0.0374)	(0.0444)
WTP		0.258		0.0650		0.0196
		(0.268)		(0.0555)		(0.0211)
lower SRI returns		-0.381*		-0.0286		0.00303
		(0.219)		(0.0464)		(0.0186)
lower SRI risk		0.221		0.0462		0.0154
		(0.234)		(0.0481)		(0.0188)
risk aversion		0.398*		0.0573		0.00980
		(0.212)		(0.0459)		(0.0185)
income		0.132		0.0586		0.0358*
		(0.230)		(0.0465)		(0.0185)
female		0.748		0.177*		0.0799*
		(0.517)		(0.106)		(0.0414)
financial literacy		0.184		0.0704		0.00526
		(0.238)		(0.0522)		(0.0209)
inv. experience		-0.210		-0.0482		-0.0292
		(0.505)		(0.108)		(0.0426)
sample Mainz		-0.199		-0.0391		-0.0524
		(0.447)		(0.0933)		(0.0381)
ESG knowledge		-0.212		0.00755		-0.0138
		(0.530)		(0.108)		(0.0418)
Constant	20.15***	20.11***	1.998***	1.955***	0.288***	0.293***
	(0.235)	(0.608)	(0.0492)	(0.120)	(0.0217)	(0.0486)
N	643	643	643	643	643	643
adj. R squared	0.194	0.195	0.208	0.208	0.185	0.189

Notes: Table 8 presents different OLS specifications showing that individuals who can use ESG Score information in their portfolio allocation 2 invest significantly more sustainably than those unable to use ESG Score information. The dependent variables are the mean ESG Score of subjects' portfolio allocation 2 (Specifications (1) - (2)), the number of sustainable stocks in portfolio allocation 2 (Specifications (3) - (4)), and an indicator that is one if an individual buys more than two sustainable stocks in portfolio allocation 2 (Specification (5) - (6)). The explanatory variable is *receive ESG Score*, which represents an indicator variable that takes the value one if the individual receives the ESG Score information in portfolio allocation 2 and zero otherwise. See Table 7 (appendix) for detailed descriptions of all variables. Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

Table 9: Sustainability preferences and the WTP for ESG Score information are correlated with altruistic preferences.

Panel A

	(1) ESG Score	(2) ESG Score	(3) WTP	(4) WTP
altruism	0.333***		181.3***	
	(0.117)		(40.80)	
impact altruism		0.305**		168.6***
		(0.125)		(44.63)
warm glow altruism		0.377***		194.4***
		(0.141)		(51.21)
Constant	20.55***	20.58***	2902.4***	2919.3***
	(0.766)	(0.767)	(271.4)	(271.5)
N	643	643	643	643
adj. R squared	0.0306	0.0291	0.107	0.105
controls	yes	yes	yes	yes

Panel B

	(1) # sustainable	(2) # sustainable	(3) # sustainable d.	(4) # sustainable d.
altruist	0.393***		0.108*	
	(0.137)		(0.0559)	
impact altruist		0.473***		0.123**
		(0.150)		(0.0611)
warm glow altruist		0.328**		0.0953
		(0.145)		(0.0593)
Constant	2.297***	2.286***	0.404***	0.401***
	(0.158)	(0.158)	(0.0657)	(0.0659)
N	643	643	643	643
adj. R squared	0.0321	0.0334	0.0228	0.0218
controls	yes	yes	yes	yes

Notes: Table 9 presents different OLS specifications demonstrating that different types of altruistic preferences are significantly correlated with sustainable investment choices. The dependent variables in Panel A are the mean ESG Score of subjects' portfolio allocation 1 (Specifications (1) - (2)) and the WTP for ESG Score information (Specifications (3) - (4)). Panel A has three explanatory variables: (i) *altruism*, quantified by the donation amount in the standard dictator game; (ii) *warm glow altruism*, represented by donations in the "money-burning" dictator game; (iii) *impact altruism*, calculated as the difference between donations in the standard and "money-burning" dictator games. The dependent variables in Panel B are the number of sustainable stocks in the baseline portfolio allocation (Specifications (1) - (2)) and an indicator variable that is one if an investor's portfolio contains more than two sustainable stocks in portfolio allocation 1. We use three variations of altruistic preferences as explanatory variables: *altruist*, an indicator variable that is one if the participant donates something in at least one of the dictator games; *impact altruist*, an indicator variable that is one if the individual donates exclusively in the standard dictator game; *warm glow altruist*, an indicator variable that is one if the participant donates in both dictator games. Furthermore, we include the same control variables in each regression as in Table 2. Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

Table 10: The randomization in the information and incentive treatment overall worked.

Panel A: Randomization information treatment

	over-performance	under-performance	Diff.	Std. Error	Obs.
ESG Score (allocation 1)	23.43	22.33	-1.11**	0.53	643
ESG Score (allocation 2)	22.15	21.73	-0.42	0.47	643
WTP	3289.26	3456.10	166.84	193.21	643
non altruist	0.15	0.16	0.01	0.03	643
impact altruist	0.34	0.38	0.04	0.04	643
warm glow altruist	0.51	0.46	-0.05	0.04	643
lower SRI returns	3.71	3.89	0.18	0.16	643
lower SRI risk	4.79	4.76	-0.03	0.14	643
risk aversion	0.47	0.46	-0.01	0.01	643
income	2.00	2.03	0.04	0.09	643
female	0.62	0.63	0.02	0.04	643
financial literacy	7.29	7.39	0.10	0.16	643
inv. experience	0.45	0.47	0.03	0.04	643
sample Mainz	0.47	0.54	0.07*	0.04	643
ESG knowledge	0.43	0.37	-0.06	0.04	643
warm glow first	0.51	0.50	-0.01	0.04	643

Panel B: Randomization incentive treatment

	positive incentive	negative incentive	Diff.	Std. Error	Obs.
ESG Score (allocation 1)	22.75	22.97	0.23	0.53	643
ESG Score (allocation 2)	22.05	21.82	-0.23	0.47	643
ESG Score (allocation 3)	20.97	20.93	-0.04	0.70	643
WTP	3452.27	3297.77	-154.50	193.14	643
non altruist	0.13	0.18	0.06**	0.03	643
impact altruist	0.35	0.37	0.02	0.04	643
warm glow altruist	0.52	0.45	-0.07*	0.04	643
lower SRI returns	3.83	3.76	-0.07	0.16	643
lower SRI risk	4.92	4.62	-0.31**	0.14	643
risk aversion	0.47	0.45	-0.02*	0.01	643
income	2.03	2.00	-0.03	0.09	643
female	0.65	0.60	-0.05	0.04	643
financial literacy	7.25	7.43	0.18	0.16	643
inv. experience	0.42	0.50	0.08**	0.04	643
sample Mainz	0.50	0.51	0.01	0.04	643
ESG knowledge	0.40	0.40	-0.00	0.04	643
warm glow first	0.50	0.50	0.01	0.04	643

Notes: Table 10 verifies that the randomization in the information treatment (Panel A) and incentive treatment (Panel B) mainly worked. To do so, mean values of all relevant variables are computed and compared for each treatment group by using t-tests. For definitions of the variables, please consider Table 7 (appendix). Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively.

Table 11: Participants' investment behavior reacts to information regarding the overperformance and the underperformance of sustainable stocks.

Panel A

	(1) # sustainable	(2) # sustainable	(3) # sustainable	(4) # sustainable
overperformance	1.256***	1.255***	1.094***	1.365***
	(0.0727)	(0.198)	(0.128)	(0.0979)
underperformance	-0.305***	0.132	-0.216*	-0.529***
	(0.0813)	(0.168)	(0.126)	(0.128)
Constant	2.810***	2.430***	2.944***	2.833***
	(0.0274)	(0.0645)	(0.0450)	(0.0401)
N	1286	200	462	624
fixed effects	yes	yes	yes	yes
sample split	-	non altruists	impact altruists	warm glow altruists

Panel B

	(1) # sustainable d.	(2) # sustainable d.	(3) # sustainable d.	(4) # sustainable d.
overperformance	0.295***	0.319***	0.264***	0.308***
	(0.0274)	(0.0687)	(0.0470)	(0.0388)
underperformance	-0.118***	0	-0.0640	-0.203***
	(0.0340)	(0.0762)	(0.0518)	(0.0536)
Constant	0.607***	0.490***	0.636***	0.622***
	(0.0110)	(0.0259)	(0.0177)	(0.0164)
N	1286	200	462	624
fixed effects	yes	yes	yes	yes
sample split	-	non altruists	impact altruists	warm glow altruists

Notes: Table 11 uses individual fixed effects regressions showing that investors' investment behavior significantly reacts to receiving risk-adjusted overperformance and underperformance information. The dependent variable *# sustainable* in Panel A, indicates the number of sustainable stocks of individual i at time $t \in \{1, 3\}$ (baseline and post-treatment portfolio allocations). The dependent variable *# sustainable d.* in Panel B, represents an indicator variable that is one if individual i at time $t \in \{1, 3\}$ has more than two sustainable shares in the portfolio (baseline and post-treatment portfolio allocations). The explanatory variable *overperformance* is an indicator variable set to one if individual i received the overperformance information at time t, and zero otherwise. Similarly, *underperformance* indicates whether individual i received underperformance information at time t. Specification (2) includes only altruistic investors, while Specification (3) is restricted to impact altruists and Specification (4) comprises exclusively warm glow altruists. Significance levels are indicated by *, **, and *** for the 10, 5, and 1 percent levels, respectively.

Table 12: Warm glow driven altruists react strongest on performance information.

	(1) Δ ESG	(2) Δ ESG	(3) Δ # sustainable	(4) Δ # sustainable	(5) Δ # sustainable d.	(6) Δ # sustainable d.
overperformance	8.733***	8.518***	1.310***	1.256***	0.328***	0.326***
	(1.063)	(0.902)	(0.179)	(0.148)	(0.0699)	(0.0577)
warm glow altruist	-1.650	-2.107**	-0.313*	-0.417**	-0.139*	-0.158**
	(1.056)	(0.963)	(0.179)	(0.164)	(0.0745)	(0.0685)
overperformance * warm glow altruist	2.903**	3.118**	0.584**	0.638***	0.183*	0.185**
	(1.404)	(1.286)	(0.241)	(0.219)	(0.0962)	(0.0878)
Constant	-2.456***	-1.999***	-0.216*	-0.112	-0.0640	-0.0449
	(0.779)	(0.648)	(0.126)	(0.102)	(0.0517)	(0.0428)
N	543	643	543	643	543	643
sample split	excl. non altruists	-	excl. non altruists	-	excl. non altruists	-

Notes: Table 12 displays an OLS regression showing that warm glow altruists react significantly stronger to performance information than impact altruists and non-altruists across various specifications of sustainability. The dependent variable is the difference between three pre- and post-treatment sustainability measures: *ESG Scores* (Specifications (1) - (2)); number of sustainable stocks (Specifications (3) - (4)); and an indicator variable that is one if an investor's portfolio consists of more than two sustainable stocks (Specifications (5) - (6)). *Overperformance* is an indicator variable that is one if a participant receives the overperformance information, while it is zero if the participant is part of the underperformance treatment. *Warm glow altruist* is an indicator variable that is one if the participant donates in both dictator games. Lastly, *overperformance * warm glow altruist* is the interaction effect of the two previously specified variables. Specifications (1), (3), and (5) exclude non-altruists, while Specifications (2), (4), and (6) include all participants. Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

Table 13: Investors' investment behavior reacts to both direct financial incentives and information on the performance of stocks.

	(1) % SRI	(2) % SRI	(3) % SRI	(4) % SRI
positive incentive	25.14***	25.60***	26.97***	28.58***
	(3.147)	(2.766)	(3.071)	(3.041)
overperformance		23.73***		16.63***
		(2.781)		(3.031)
Constant	45.33***	36.80***	44.13***	53.00***
	(2.207)	(13.15)	(2.191)	(16.59)
N	311	300	332	318
individual controls	no	yes	no	yes
split	no study	no study	study	study
adj. R squared	0.169	0.404	0.187	0.292

Notes: Table 13 uses OLS regressions to show that participants' investment behavior reacts to direct financial incentives as well as to performance information. The dependent variable is the share of investments in the sustainable portfolio in allocation 4. The explanatory variable *positive incentive* is an indicator variable taking the value of one if an investor is part of the positive incentive treatment group and zero otherwise. *Overperformance* is an indicator variable that is one if an investor is part of the overperformance treatment and zero otherwise. Specifications (1) and (2) focus on participants without the possibility to read an article on the moral consequences, while Specifications (3) and (4) exclusively include participants that read a story on the moral consequences of SRI. Lastly, Specifications (1) and (3) do not include controls, while Specifications (2) and (4) include our standard set of control variables (see e.g., Table 2 for reference on the controls). Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

Table 14: Investors' share of sustainable investments varies with different levels of incentives.

	(1) %SRI	(2) %SRI
incentive of -20%	-27.88***	-25.90***
	(1.549)	(1.547)
incentive of -10%	-20.41***	-19.13***
	(1.549)	(1.547)
incentive of +10%	8.167***	9.867***
	(1.549)	(1.547)
incentive of +20%	15.41***	16.32***
	(1.549)	(1.547)
Constant	65.23***	63.81***
	(1.096)	(1.094)
N	1555	1660
fixed effects	yes	yes
sample split	no study	study

Notes: Table 14 uses individual fixed effect regressions and shows that the share of investments in the sustainable portfolio changes for different incentive levels to invest in the high sustainability portfolio. The dependent variable is the share of sustainable investments in a given portfolio allocation at time $t \in \{4, 5, 6, 7, 8\}$. The explanatory variables are indicator variables that are one if the level of incentive to invest sustainably at time t is $z \in \{-20\%, -10\%, +10\%, +20\%\}$ and zero otherwise. For example, the variable *incentive of -20%* indicates that in the portfolio allocation in question, the proctor reduces the amount of investments in the sustainable portfolio by 20%. Note, the incentive of 0% is the baseline category. Specification (1) exclusively focuses on investors without the opportunity to select a story, while Specification (2) considers only those subjects reading one of the stories on the moral consequences of SRI. Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

Table 15: (Perceived) incentives skew information acquisition. This is driven by warm glow altruists

	(1) moral bad	(2) moral bad	(3) moral bad	(4) moral bad
clear positive incentives	-0.156*	-0.0170		
	(0.0793)	(0.108)		
unclear incentives	-0.0957	-0.0222		
	(0.0656)	(0.0896)		
warm glow altruist		0.0965		0.0661
		(0.108)		(0.0770)
clear positive incentives * warm glow altruist		-0.296*		
		(0.157)		
unclear incentives * warm glow altruist		-0.163		
		(0.131)		
positive incentive			-0.0630	0.0608
			(0.0543)	(0.0746)
positive incentive * warm glow altruist				-0.251**
				(0.108)
Constant	0.662***	0.622***	0.607***	0.578***
	(0.0541)	(0.0729)	(0.0384)	(0.0524)
N	332	332	332	332
adj. R squared	0.00570	0.0106	0.00104	0.0151

Notes: Table 15 uses OLS regressions and presents evidence that investors select the article on the moral behavior of SRI most in line with their (perceived) financial incentives. This effect is driven by warm glow altruists. The dependent variable *moral bad* refers to a dummy variable that is one if an individual selects the article *"Why investing in sustainable stocks makes no sense from a moral point of view"* and zero if the individual chooses *"Why investing in sustainable stocks makes sense from a moral point of view"*. *Clear positive incentives* is an indicator variable equal to one if the investor is subject to the overperformance treatment group and, in a later stage of the experiment, is part of the positive incentive treatment group. Contrarily, *unclear incentives* is an indicator variable equal to one if the investor is subject to the overperformance treatment group and, in a later stage of the experiment, is part of the negative incentive treatment group or is subject to the underperformance treatment group and later is part of the positive incentive treatment group. *Warm glow altruist* is an indicator variable that is one if the participant donates in both dictator games. Lastly, *positive incentive* is an indicator variable equal to one if the investor is part of the positive incentive treatment group. In line with our pre-analysis plan, we exclude all individuals of the non-study group in this analysis. Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

Table 16: Relation between skewed information acquisition and SRI.

	(1) % SRI
study group	-1.200
	(3.111)
positive incentive	25.14***
	(3.160)
study * positive incentive	1.824
	(4.398)
Constant	45.33***
	(2.217)
N	643
adj. R squared	0.177
controls	no

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Table 16 shows an OLS specification analyzing the transmission of the skewed information acquisition in sustainable investments (pre-registered specification). The dependent variable *% SRI* indicates the share of sustainable investments in portfolio allocation 4. *Study group* is a dummy variable indicating whether a subject gets the opportunity to read an article about the moral perspectives of SRI. Lastly, *positive incentive* is an indicator variable equal to one if the investor is part of the positive incentive treatment group. *Study * positive incentive* represents the interaction term of the former two variables. Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

Table 17: (Perceived) incentives and information acquisition (full sample)

Panel A: Characteristics moral bad story and matched control group

	moral bad story	matched control	Diff.	Std. Error	Obs.
positive incentive	0.48	0.48	0.00	0.05	382
overperformance	0.45	0.45	0.00	0.05	382
# sustainable (allocation 1)	2.73	2.75	0.03	0.12	382
# sustainable (allocation 2)	2.24	2.38	0.14	0.12	382
# sustainable (allocation 3)	3.20	3.32	0.13	0.15	382
non altruist	0.18	0.18	0.00	0.04	382
impact altruist	0.37	0.37	0.00	0.05	382
warm glow altruist	0.29	0.29	0.00	0.05	382
partly warm glow altruist	0.16	0.16	0.00	0.04	382
warm glow first	0.52	0.47	-0.06	0.05	382
lower SRI returns	3.70	3.92	0.22	0.21	382
lower SRI risk	4.66	4.96	0.30	0.18	382
risk aversion	0.46	0.47	0.01	0.01	382
income	1.96	2.15	0.19	0.12	382
female	0.58	0.57	-0.01	0.05	382
financial literacy	7.65	7.81	0.16	0.18	382
inv. experience	0.48	0.54	0.06	0.05	382
sample Mainz	0.51	0.48	-0.03	0.05	382
ESG knowledge	0.43	0.46	0.03	0.05	382
WTP	3210.48	3117.86	-92.61	231.92	382

Panel B: Characteristics moral good story and matched control group

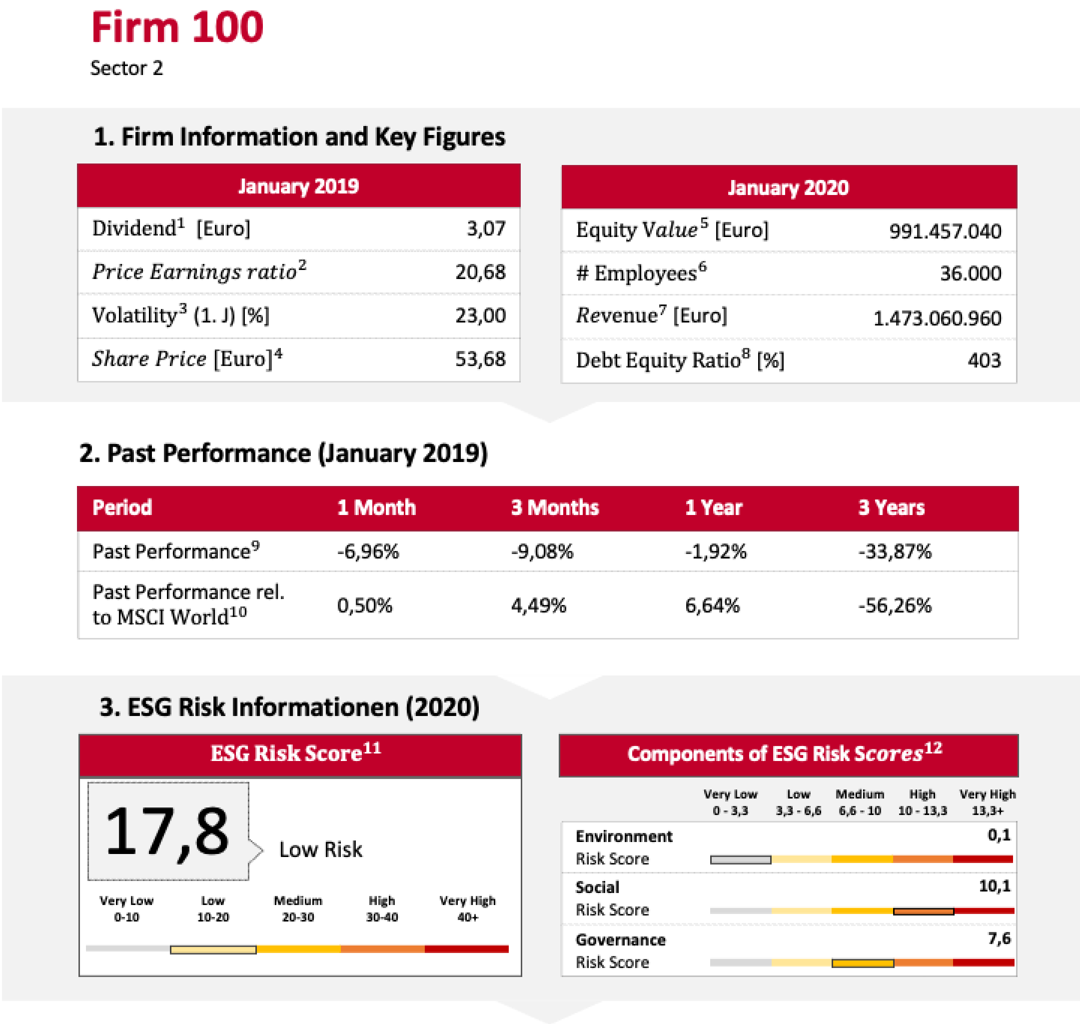
	moral good story	matched control	Diff.	Std. Error	Obs.
positive incentive	0.55	0.55	0.00	0.06	282
overperformance	0.53	0.53	0.00	0.06	282
# sustainable (allocation 1)	2.81	2.98	0.17	0.14	282
# sustainable (allocation 2)	2.33	2.32	-0.01	0.13	282
# sustainable (allocation 3)	3.16	3.36	0.20	0.17	282
non altruist	0.12	0.12	0.00	0.04	282
impact altruist	0.36	0.36	0.00	0.06	282
warm glow altruist	0.28	0.28	0.00	0.05	282
partly warm glow altruist	0.24	0.24	0.00	0.05	282
warm glow first	0.47	0.51	0.04	0.06	282
lower SRI returns	3.84	3.78	-0.06	0.22	282
lower SRI risk	4.77	4.85	0.08	0.19	282
risk aversion	0.49	0.47	-0.02	0.01	282
income	1.97	2.06	0.09	0.14	282
female	0.68	0.67	-0.01	0.06	282
financial literacy	6.94	7.35	0.40*	0.23	282
inv. experience	0.45	0.48	0.02	0.06	282
sample Mainz	0.48	0.52	0.04	0.06	282
ESG knowledge	0.43	0.40	-0.04	0.06	282
WTP	3538.45	3113.29	-425.16	270.95	282

Notes: Table 17 reports the full sample and shows that the mean values of a wide array of controls are not significantly different between those selecting the moral bad (good) story and their matched control partner. For every participant who reads one of the two articles, we search for a perfect match among the participants who did not have the chance to read an article. We use exact matches for the information and incentive treatment as well as for the altruistic type. Additionally, we look for the nearest neighbors in expected ESG return, gender, risk attitude, and financial literacy. Panel A reports mean values and t-tests of subjects choosing the moral bad story and their matched partners, while Panel B reports this information for those selecting the moral good story. *, **, and *** denote significance at the 10, 5, and 1 percent levels.

Table 18: Reading articles on the moral consequences of sustainable investments influences investment decisions (full sample)

	(1) % SRI	(2) % SRI	(3) % SRI	(4) % SRI
reading moral bad	-13.55***	-4.247	-13.54**	-15.36***
	(3.805)	(11.83)	(6.572)	(4.954)
Constant	6.266**	-6.353	8.127	7.904**
	(2.886)	(9.705)	(4.999)	(3.644)
N	332	52	121	159
sample split	-	non altruist	impact altruist	warm glow altruist
matching	nnm	nnm	nnm	nnm

Notes: Table 18 includes the full sample of individuals reading an article. The OLS regression demonstrates that reading the moral bad article significantly reduces the share of sustainable investments compared to reading the moral good article. The dependent variable $ATET$ is the average treatment effect on the treated of reading the moral bad (good) article. We calculate the ATET for each participant in the study group by subtracting the SRI share of the matched control (nearest neighbor matching) from the SRI share of the respective participant in portfolio allocation 4. The explanatory variable *reading moral bad* is an indicator variable set to one if the participant reads the moral bad research article and zero if the participant reads the moral good article. Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.



Firm 100

Sector 2

1. Firm Information and Key Figures

January 2019	
Dividend[1] [Euro]	3,07
Price Earnings ratio[2]	20,68
Volatility[3] (1. J) [%]	23,00
Share Price [Euro][4]	53,68

January 2020	
Equity *Value*[5] [Euro]	991.457.040
# Employees[6]	36.000
Revenue[7] [Euro]	1.473.060.960
Debt Equity Ratio[8] [%]	403

2. Past Performance (January 2019)

Period	1 Month	3 Months	1 Year	3 Years
Past Performance[9]	-6,96%	-9,08%	-1,92%	-33,87%
Past Performance rel. to MSCI World[10]	0,50%	4,49%	6,64%	-56,26%

3. ESG Risk Informationen (2020)

ESG Risk Score[11]

17,8 Low Risk

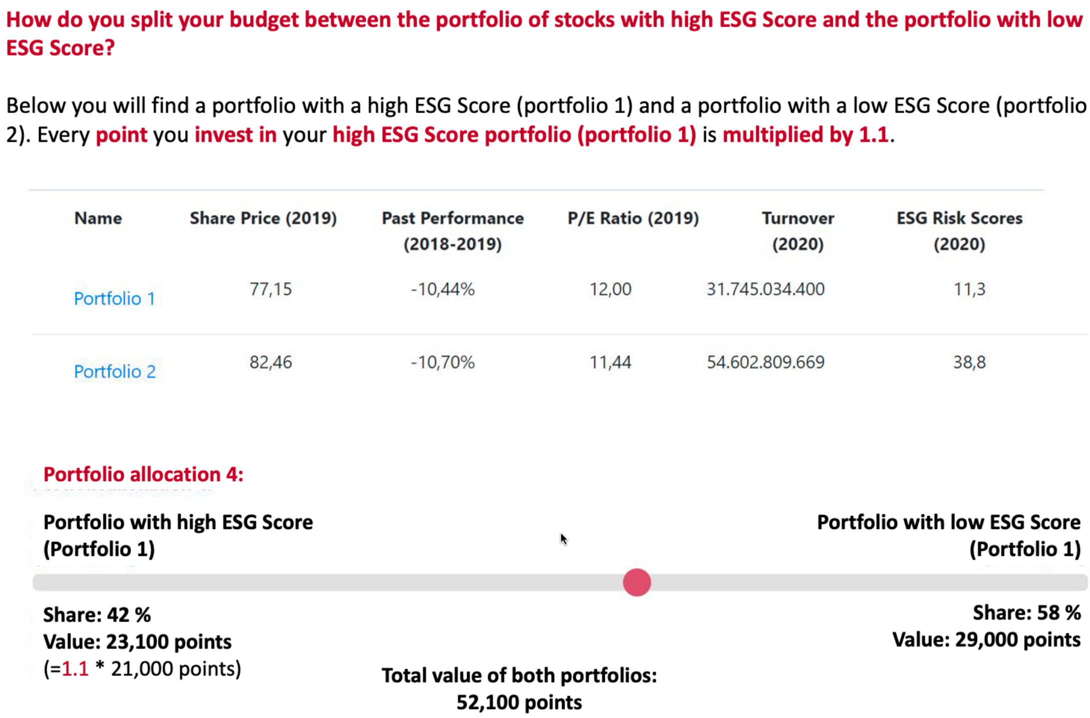
Very Low 0-10 | Low 10-20 | Medium 20-30 | High 30-40 | Very High 40+

Components of ESG Risk Scores[12]

	Very Low 0 - 3,3	Low 3,3 - 6,6	Medium 6,6 - 10	High 10 - 13,3	Very High 13,3+
Environment Risk Score: 0,1					
Social Risk Score: 10,1					
Governance Risk Score: 7,6					

Figure 5: Fact sheet of stock 100

Notes: This figure illustrates a translated version of a typical fact sheet from the stock market game used in the experiment. Participants could view this fact sheet by clicking on a stock's name within the trading interface. The fact sheet includes data from 2019, such as dividend yield, price-to-earnings (PE) ratio, volatility, share price, and past performance. Additionally, it features 2020 information like equity value, number of employees, turnover, and debt ratio. It also provides details about the ESG Risk Score, enhancing participants' ability to make informed investment decisions based on both financial metrics and sustainability factors.



Name	Share Price (2019)	Past Performance (2018-2019)	P/E Ratio (2019)	Turnover (2020)	ESG Risk Scores (2020)
Portfolio 1	77,15	-10,44%	12,00	31.745.034.400	11,3
Portfolio 2	82,46	-10,70%	11,44	54.602.809.669	38,8

Figure 6: Allocation of endowment between a portfolio with high and a portfolio with low ESG Score

Notes: This figure displays portfolio allocation 4 (positive incentive treatment group). Here, subjects get to decide between two portfolios where portfolio 1 has a high ESG Score and portfolio 2 a low one. By moving the slider, participants can choose the share of their endowment that they want to invest in the portfolio with a high or a low ESG Score, respectively. By clicking on the names of the portfolios, subjects receive a fact sheet for each portfolio similar to Figure 5.

Chapter 3

3 Shifting Grounds: The Influence of Motivated Beliefs on Investor's Convictions about Sustainability

Fabian Braun[ab] Andrej Gill[ac] Florian Hett[ad]

August 2024

Abstract

How do retail investors define sustainable investments? Given the political efforts to attract private capital to facilitate Europe's transition to a more sustainable economy, this question is crucial. Our incentivized online experiment with 652 German residents provides causal evidence that investors' sustainability beliefs are influenced by motivated beliefs. We exogenously vary participants' return expectations and their knowledge about the sustainability attributes of the investments they selected earlier in the experiment. The results show that the interaction of these two factors shapes beliefs about what an appropriate definition of sustainable investing should include. Our findings have implications for researchers and policymakers alike, highlighting the importance of considering motivated beliefs in the development of regulatory frameworks, and economic models.

Keywords: Sustainable Finance, Motivated Reasoning, Sustainability Convictions, MiFID II

[a]Comments from audiences at Goethe University Frankfurt, Leibniz Institute for Financial Research SAFE, and Johannes Gutenberg University Mainz are gratefully acknowledged. The joint ethics committee of the Faculty of Economics and Business Administration of Goethe University Frankfurt and the Gutenberg School of Management & Economics of the Faculty of Law, Management and Economics of Johannes Gutenberg University Mainz granted the project exempt status. The experiment is pre-registered under AEARCTR-0012566 and if not stated otherwise, we follow a pre-registered analysis plan. We thank Max Herbert for outstanding research assistance. We gratefully acknowledge research support from the Leibniz Institute for Financial Research SAFE. The authors declare no conflict of interest.

[b]Johannes Gutenberg University Mainz; Jakob-Welder-Weg 9, 55128 Mainz, Germany. Phone: +49 6131 3927901. Email: braun@uni-mainz.de

[c]Johannes Gutenberg University Mainz; Jakob-Welder-Weg 9, 55128 Mainz, Germany. Phone: +49 6131 3922979. E-Mail: gill@uni-mainz.de

[d]Johannes Gutenberg University Mainz; Johann-Joachim-Becher-Weg 31, 55128 Mainz, Germany. Phone: +49 6131 39-20567. E-Mail: florian.hett@uni-mainz.de

3.1 Introduction

The European Green Deal initiative aims to embed sustainability at the heart of the financial system, facilitating Europe's transition to a more sustainable and resilient economy (Commission, 2019a). In line with this, the August 2022 update to the Markets in Financial Instruments Directive II (MiFID II) aims to incorporate client sustainability preferences into the Investor Suitability Assessment (Commission, 2021a).[25] The regulation requires investment advisors to inquire whether their clients wish to include sustainable investment products in their portfolios, and if so, to what extent.[26] This process is designed to provide clients with definitions of sustainability that meet their preferences. However, anecdotal evidence suggests that, in practice, advisors may influence clients to avoid certain sustainability preferences in order to maintain a broader universe of investment opportunities, potentially leading to washed-out sustainability preferences. The literature shows that investors often have a preference to invest sustainably (Barreda-Tarrazona *et al.*, 2011; Hartzmark and Sussman, 2019; Bauer *et al.*, 2021). This preference is influenced by risk and return expectations on sustainable investments (Døskeland and Pedersen, 2016; Giglio *et al.*, 2023; Braun *et al.*, 2024b). However, the market for sustainable investments is highly ambiguous,[27] which facilitates the emergence of motivated beliefs (Gino *et al.*, 2016). This can enable strategic information acquisition, which in turn changes subsequent sustainable investment decisions (Braun *et al.*, 2024b). Our experiment aims to test the role of motivated beliefs on investors' perceptions of sustainability. We examine how investors' evaluations of what sustainability should entail are influenced by their return expectations and their knowledge of the sustainability characteristics of their previously selected investments in the experiment, both of which may significantly influence motivated beliefs.

To answer this question, we conduct an incentivized online experiment with 652 German residents. Initially, participants are informed about three different sustainability concepts (*low emission intensity*, *exclusion of controversial companies*, and *alignment with the 2°C goal*). We then randomly assign them to either the *positive treatment* or the *negative treatment* group, where they learn about the relationship between these sustainability definitions and the expected one-year financial returns of 14 real Exchange-Traded Funds (ETFs). Individuals in the *positive treatment* observe a positive correlation between the expected returns of ETFs comply-

[25]The final guidelines on certain aspects of the MiFID II suitability requirements were published in April 2023 on the European Securities and Markets Authority website and came into force from October 2023 (ESMA, 2023).

[26]The amendment to the Investor Suitability Assessment (ESMA, 2023) involves asking clients whether they wish to include financial products that: (i) allocate a specified minimum percentage to environmentally sustainable investments as defined in the EU Taxonomy Regulation; (ii) allocate a minimum percentage to sustainable investments as defined in the Sustainable Finance Disclosure Regulation (SFDR); or (iii) take into account the principal adverse impacts on sustainability factors. In addition, clients specify the minimum proportion of sustainable products in the total investment.

[27]There exists no universal definition on what sustainable investments must entail (Filippini *et al.*, 2024). Each region develops their own definition of sustainability. In addition, the commonly used ESG Scores that are supposed to indicate the sustainability of companies vary widely from one rating agency to the next (Dimson *et al.*, 2020; Billio *et al.*, 2021; Berg *et al.*, 2022). Last, it is still unclear whether sustainable stocks outperform or underperform conventional stocks (Eccles *et al.*, 2014; Friede *et al.*, 2015; Pástor *et al.*, 2020; Pastor *et al.*, 2021; Avramov *et al.*, 2022; Bolton and Kacperczyk, 2023; Latino, 2023; Bolton and Kacperczyk, 2024).

ing with the definition *alignment with the 2°C goal*, while individuals in the *negative treatment* observe a negative correlation. Participants then make incentivized decisions on how much of their available funds they invest sustainably. In the next step, they are either randomly assigned a sustainable ETF (*passive treatment group)* or allowed to choose their preferred sustainability definition, which then matches them with a corresponding ETF (*active treatment group*). Finally, we present participants with arguments for and against each sustainability definition and assess their evaluation of these definitions for classifying sustainable investments. In this context, we aim to determine whether investors' assessment of sustainability definitions remains consistent, regardless of return expectations or sustainability attributes of existing portfolio holdings. We interpret any correlation between our exogenously varied factors and the evaluation of sustainability definitions as an indication of the presence of motivated beliefs.

Our research documents three sets of results. First, we provide evidence that return expectations play a crucial role in sustainable investment choices. We demonstrate that changes in return expectations alter the choice of sustainable investments among different sustainable options. The selection of ETFs aligned with the 2°C goal sustainability definition in the *positive treatment* is 6.5 times more likely than in the *negative treatment*, highlighting a strong positive correlation between expected financial returns and the preference for ETFs adhering to the 2°C goal definition. Second, our findings highlight that shifts in return expectations have a significant impact on people's perceptions of what characteristics should define sustainability. In particular, participants in the *positive treatment* exhibit a 61% greater preference for *alignment with the 2°C goal* as their preferred sustainability definition compared to those in the negative return treatment. This observation challenges the premise that sustainability assessments should be based on a set of clear and objective criteria for the performance of sustainability attributes. Rather, our findings indicate that participants' convictions about sustainability are influenced by considerations related to returns. Third, our findings suggest that participants' sustainability convictions are mostly influenced by motivated beliefs induced through the interplay of return expectations and existing portfolio holdings in the experiment. According to Taleb (2018), someone with *skin in the game* possesses a vested interest in the outcome of an event and, crucially, stands to lose something if it doesn't occur. In our case, participants form motivated beliefs about what a suitable sustainability definition should entail if they possess assets with certain sustainability characteristics. Our analysis documents that the effects of motivated beliefs on the selection of the best sustainability definition, arising from different return expectations, are about five times larger in the *active treatment* than in the *passive treatment* group. Notably, this discrepancy is predominantly attributed to participants who actually possess investments aligned with the 2°C goal and therefore have a vested interest that their portfolio holdings are considered sustainable.

Our paper bridges two key research areas and connects to recent regulatory developments in the area of sustainable finance. First, we contribute to the expanding body of literature on sustainable finance which tries to answer the question of what drives sustainable investments (Døskeland and Pedersen, 2016; Riedl and Smeets, 2017; Heeb *et al.*, 2023; Giglio *et al.*, 2023; Braun *et al.*, 2024b). Among others, work by Barreda-Tarrazona *et al.* (2011); Hartzmark and Sussman (2019); Bauer *et al.* (2021) shows that investors integrate sustainability preferences into their investment decisions. Døskeland and Pedersen (2016) introduce a wealth information treatment and demonstrate that investors significantly react to financial cues. This is in line with Giglio *et al.* (2023), who show that investors choose sustainable investments only when they anticipate these investments to outperform, regardless of other investment motives. Moreover, our findings add to the work of Riedl and Smeets (2017); Bauer *et al.* (2021); Gutsche *et al.* (2023), who suggest that sustainable investments are predominantly driven by non-financial motivations, with investors potentially willing to forgo higher returns for sustainable options. Altruistic individuals may even reduce their motivation to invest sustainable, if they expect higher financial returns (Brodback *et al.*, 2019). Our research differs from previous studies by examining how investors choose among different sustainable investment options.

Second, our study extends the growing literature on motivated beliefs (Epley and Gilovich, 2016; Bénabou and Tirole, 2016; Eyting, 2022; Amelio and Zimmermann, 2023)[28] by linking these concepts to sustainability. We investigate whether investors exhibit self-serving behavior by preferring sustainability definitions that resonate with their return expectations and existing portfolio holdings. According to Epley and Gilovich (2016) individuals tend to reason their way to conclusions that favor them. People´s motivation influences how arguments are gathered and processed or how past experiences are recalled, resulting in beliefs that seem to be objective but truly are biased in one way or another. Following Gino *et al.* (2016) people are often willing to take a moral act (sustainable investments in this experiment) that imposes personal costs when confronted with a clear-cut choice but such decisions often seem to be dramatically influenced by the specific contexts in which they occur. When the context provides sufficient flexibility to allow plausible justification that one can both act egoistically while remaining moral, people seize on such opportunities to prioritize self-interest at the expense of morality. In the context of this experiment, the presence of various sustainability definitions, each with ample justification for being considered sustainable, leads participants to choose the one that offers them the greatest benefit. Note, this is a simplified experiment, in a real-world decision this problem is way more complicated which amplifies the potential for motivated beliefs. In an application of these principles Braun *et al.* (2024b) show that incentives change the acquisition of information on the moral consequences of sustainable investments and this subsequently alters investment behavior. We differ from the existing literature in a way that we show how motivated beliefs change the assessment of sustainability definitions with real-world consequences.

[28]See Epley and Gilovich (2016) for an overview on the mechanisms behind motivated beliefs.

Last, our findings hold significant policy implications for both the development of future regulations and the refinement of existing ones. We urge for more precise definitions within the MiFID II framework (Commission, 2021a), suggesting a move towards simplifying sustainability definitions to the greatest extent possible, ideally converging on a singular, universal definition characterized by clear, measurable criteria. Currently, there is an ongoing discussion around the implementation regarding clear labeling of sustainable products. For instance, the Sustainable Finance Advisory Board of the Federal Government in Germany (2024) proposes the inclusion of a straightforward scale from A to F for retail finance products to enhance their comprehensibility. As outlined by Gino *et al.* (2016), motivated beliefs thrive in environments of ambiguity. The present regulatory landscape, with three possible definitions and an extensive list of economic activities and considerations for sustainable investments, poses a challenge for retail investors and investment advisors. This complexity not only fosters motivated reasoning during the assessment of sustainable assets but also obstructs the genuine expression of sustainability preferences. Moreover, the existing ambiguity within these regulations could potentially be exploited by advisors to sway clients' sustainability preferences, thereby promoting higher-priced sustainable products (Laudi *et al.*, 2023). This underscores the need for regulatory clarity to safeguard investors' interests and ensure that sustainability preferences are authentically represented and effectively integrated into investment decisions.

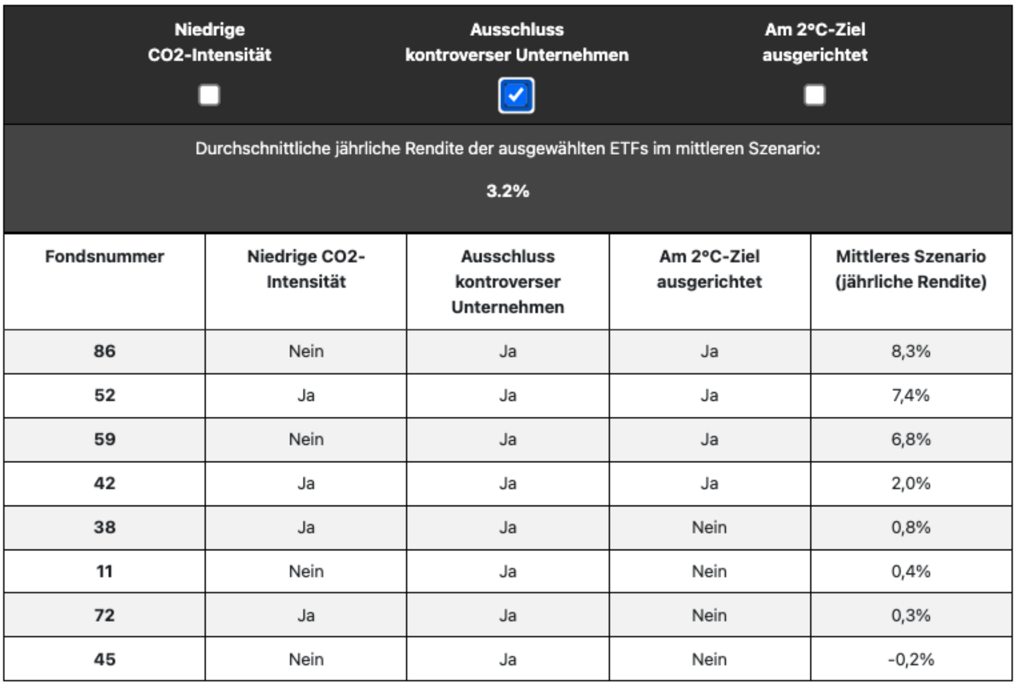
The remainder of the paper is structured as follows: In section 3.2 we provide details on the experimental design, key variables, and data collection. Section 3.3 presents the main results. Section 3.4 discusses the results and concludes.

3.2 Experimental Design, Key Variables, and Data Collection

3.2.1 Design of the Experiment and Important Variables

Our experiment comprises five stages: (i) a baseline stage to elicit participants' initial return beliefs; (ii) a first treatment stage, where participants learn about the correlation between three sustainability definitions and potential return scenarios; (iii) a second treatment stage, where investors make an active (passive) sustainable investment choice; (iv) the main stage, involving the evaluation of different sustainability definitions; and (v) a final stage for eliciting posterior beliefs and gathering a set of background characteristics. A translated version of the experimental instructions (German to English) is available upon request.

Baseline Stage: Prior Return Beliefs Participants learn about three distinct sustainability definitions that are used in standard iShares fact sheets: (i) *Low emission intensity* indicates that companies in an ETF emit less than 80 tons of CO_2 per \$1 million in revenue. (ii) *exclusion of controversial companies* omits firms from an ETF that have revenues tied to controversial weapons, unconventional oil and gas exploration, coal, or tobacco. This sustainability definition also excludes ETFs investing more than 1.5% of their portfolio in companies with UN Global Compact violations. (iii) *alignment with the 2°C goal* suggests that if global emissions followed the trend of companies in the portfolio, the temperature rise could be limited to between 1.5°C and 2°C. These sustainability definitions allow us to leverage real-world information and provide a more precise examination of sustainability concepts than the often ambiguous ESG Scores, which can vary widely between providers (Dimson *et al.*, 2020; Berg *et al.*, 2022; Billio *et al.*, 2024). After educating participants about these definitions, we follow Giglio *et al.* (2021) and Laudenbach *et al.* (2021) to elicit respondents' point estimates of beliefs about the 12-month return of ETFs that (i) have a *low emission intensity*, (ii) *exclude controversial companies*, (iii) are *aligned with the 2°C goal*, or (iv) do *not follow any sustainability definition.*



Fondsnummer	Niedrige CO2-Intensität	Ausschluss kontroverser Unternehmen	Am 2°C-Ziel ausgerichtet	Mittleres Szenario (jährliche Rendite)
86	Nein	Ja	Ja	8,3%
52	Ja	Ja	Ja	7,4%
59	Nein	Ja	Ja	6,8%
42	Ja	Ja	Ja	2,0%
38	Ja	Ja	Nein	0,8%
11	Nein	Ja	Nein	0,4%
72	Ja	Ja	Nein	0,3%
45	Nein	Ja	Nein	-0,2%

Figure 7: ETF selection interface

Notes: This Figure displays the interactive user interface that shows the list of ETFs, indicators for every sustainability definition, and the expected median returns (*positive treatment*).

Return Treatment Stage: Correlation between Sustainability Definitions and Returns In the return treatment, we exogenously vary beliefs about the returns of ETFs in line with the 2°C goal definition. To do so participants are presented with an interactive user interface that shows a list of ETFs (see Figure 7). For every ETF we display the expected return over the following 12 months in the median scenario according to the EU regulation on Packaged Retail and Insurance-based Investment Products (PRIIPs)[29] and indicators showing compliance with the sustainability definitions explained in the baseline stage. Furthermore, we summarize the average return of all displayed ETFs at the top of the table. Participants can interactively select

[29]The EU regulation on PRIIPs mandates capital management companies to provide a fact sheet stating the expected return in different scenarios. The median scenario indicates the median annual return over a holding period of one year for a (comparable) investment product over the past 10 years.

different sustainability definitions, which filters the displayed ETFs accordingly. As a result, selecting sustainability definitions alters both the number and average return of the displayed funds. To shift participants' return beliefs, we randomly assign participants equally between the *positive treatment* and *negative treatment* group. Those in the *positive treatment* group observe 14 ETFs with a positive correlation between expected returns and the *alignment with the 2°C goal* sustainability definition, while those in the *negative treatment* group see 14 ETFs with a negative correlation. One may discern the sustainability-return relation by comparing the average return for each ETF or by utilizing the interactive display to observe the displayed average return for each sustainability combination. After learning about the correlation, participants decide how to allocate 50 euros between a sustainable ETF, randomly selected from a pre-defined universe of ETFs in the subsequent step, and a non-sustainable ETF.

Choice Treatment Stage: Knowledge on the Sustainability Attributes of the Sustainable Investment In the *choice treatment*, we exogenously vary investors' knowledge of the sustainability attributes of their investments within the experiment. Participants are randomly assigned to one of two groups. Those in the *active treatment group* are aware of the amount and the attributes of their sustainable investment, whereas those of the *passive treatment group* are only aware of the amount of their sustainable investment, but do not know the exact sustainability indicators of their investment. Both groups view a screen displaying the same ETFs and the same sustainability-return relation as on the preceding page. The key distinction is that in the *active treatment group*, participants may select up to three sustainability definitions for their sustainable investment, whereas in the *passive treatment group*, the sustainable ETF is randomly determined without subjects being able to specify the sustainability definitions. It is noteworthy that participants in the *active treatment group* are not provided with any supplementary information beyond that which is initially presented; they interact with the identical interface that displays the ETFs and their sustainability-return correlations as on the previous page. Similarly, those in the *passive treatment group* are also revisited with the comprehensive list of ETFs and their respective sustainability-return correlations. In the final phase of the *choice treatment* stage, a sustainable ETF (that meets the specified criteria) is randomly assigned.

The *passive treatment group* serves as a baseline, reflecting that investors might authorize advisors to make sustainable investments without being aware of the exact sustainability attributes of their investments. Furthermore, it mimics investors who previously bought sustainable investments but do not know the sustainability attributes of their investments. Therefore, we do not inform participants in the *passive treatment group* about the sustainability definition of their investment until after the experiment. Conversely, participants in the *active treatment group* specify their sustainability preferences before being randomly allocated a fund matching these definitions, thereby becoming aware of the attributes of their ETF. According to Taleb (2018), someone with *skin in the game* possesses an interest in the outcome of an event (in this context, the consideration of investments as sustainable) and, crucially, stands to lose something if it doesn't occur (the self-image of being a good person, as discussed by Rosenberg (1965)). Therefore, this setup explores whether investors favor certain definitions in the main stage because they own an ETF meeting those criteria. Note that our study employs a 2x2 design. Consequently, all participants are initially assigned to either the *positive treatment group* or the *negative treatment group*. Subsequently, the same participants proceed and participate in either the *active treatment* or the *passive treatment*.

Main Stage: Evaluation of Different Sustainability Definitions In the main stage, where we gather our key outcomes, we present participants with arguments for and against each sustainability definition and ask them to evaluate each criterion (*"How convincing do you find the criterion 'sustainability$_i$' for classifying sustainable investments?"*) on a scale from one (not at all convincing) to five (fully convincing). Based on these responses, we calculate the variable *2°C assessment*, which reflects whether participants perceive *alignment with the 2°C goal* as more persuasive than the other sustainability definitions. Specifically, *2°C assessment* represents the difference between the evaluation of the *alignment with the 2°C goal* definition and the average rating of the other two definitions. To deepen our insight into the rationale behind respondents' decisions, participants provide explanations for their choices in open-text fields. Subsequently, participants are asked to envision a future regulatory framework where a single definition governs sustainable investment evaluations. The response to *"Please select the criterion that you consider to be the most important for evaluating sustainable investments."* is used to establish our primary outcome (*2°C best definition*). This indicator variable is set to one if a participant identifies *alignment with the 2°C goal* as the optimal sustainability definition, and zero otherwise.

Final Stage: Elicitation of Posterior Beliefs and Demographics After the sustainability assessment, we elicit participants' posterior risk and return expectations for ETFs complying with a particular sustainability definition versus an ETF not adhering to any definition. For instance, we ask participants to rate their agreement with the statement, *"ETFs aligned with the 2°C target will generate a higher return over the next 12 months than ETFs not aligned with the 2°C target"*. We also gather information on investment experience and attitudes towards sustainable investments, followed by data on education, occupation, and income.

3.2.2 Data

Data Collection, and Incentive Compatibility Data collection occurred between November 30th and December 15th, 2023, in an oTree-programmed (Chen *et al.*, 2016) online experiment. We recruit participants currently residing in Germany via Prolific, an online platform known for high data quality (Peer *et al.*, 2021) and pay a participation fee of 3 euros. To ensure incentive compatibility, we randomly select every 25th participant to implement their investment decision on the real stock market with 50 euros.[30] This decision involves (i) the allocation of 50 euros between a sustainable and a non-sustainable ETF, and (ii) the selection of sustainability definitions for the sustainable fund in the *active treatment group*. In December 2024, these participants receive a bonus payment equivalent to the current value of their ETFs. Participants are aware that their decisions impact the financing of real firms and their final payoff.

Sample Definition In total, we collected 683 valid responses.[31] We pre-registered three steps to screen out participants who might be subject to experimenter demand effects, likely did not take the experiment seriously or quickly "clicked" through the survey: participants that (i) identify the research question, (ii) fail both attention checks or (iii) disproportionately quickly finish the study. First, we screen all the answers to our question *"Please describe in one or two sentences what you think is the main purpose of this study?"*. None of the participants identifies the research question correctly. In the next step, we exclude 29 participants who fail both attention checks. Last, we drop two participants who finish the study faster than a third of the median time (6 minutes 56 seconds). Our results are robust to different cutoffs of the completion time and including all excluded participants. The median time spent in the experiment is about 21 minutes, which leads to median hourly earnings including the bonus payment of 14.30 euros per participant[32] and is considerably more than the recommended pay by Prolific of 10.50 euros. We offer these rather high incentives to ensure that the trades in the experiment accurately reflect the choices participants would make in the real stock market.

Sample Characteristics Our final sample consists of 652 participants. Table 19 provides summary statistics for all variables. The majority of our participants is male (71%), with 47% working full-time, and the median age being 28 years.[33] The sample exhibits a relatively high level of education, with 56% holding a higher education degree. Notably, 62% of participants own stocks, bonds, or funds/ETFs before participating in the study, compared to 18% in the general German population participating in the stock market.[34] On average, participants possess more than three years of investment experience.

[30]Following Charness *et al.* (2016) this approach is at least as effective as paying all participants and can further reduce transaction costs.

[31]We pre-registered 800 participants; however, only 683 had participated by the latest date feasible for settling accounts with the accounting department before year's end. The study initially drew significant interest, with about 165 participants daily. However, the number then fell to around 20 per day after four days. To boost participation, we took two measures five days before the study's end: (i) we relaxed the 50/50 male-to-female ratio requirement, and (ii) we sent a second invitation to all potential participants. These efforts led to a temporary spike in engagement, but after two days, participation dwindled to roughly 20 per day once more.

[32]Bonus payments are set for December 2024, with an investment of 50 euros made for every 25th participant in December 2023, averaging 2 euros per participant. If the value of the ETFs purchased in December 2024 surpasses their initial investment value from December 2023, the bonus payments will adjust upwards correspondingly, and the reverse will apply if the value decreases.

[33]For one participant we had no data on age. We impute this person's age using the median age of 28. All results hold excluding this individual.

[34]Only a minor fraction of Germans exclusively invests in bonds (Kantar, 2023).

Table 19: Summary statistics of the main variables for the prolific sample

	Mean	SD	Min	Max	N
returns 2°C	7.58	10.68	-5.26	70.00	652
returns exclusion	7.32	11.58	-10.00	70.00	652
returns CO2	9.48	16.09	-6.32	120.00	652
returns non-SRI	6.19	13.44	-50.00	75.80	652
return uncertainty	2.38	1.01	1.00	5.00	652
investment experience	3.04	4.69	0.00	40.00	652
own SRI	0.26	0.44	0.00	1.00	652
plans to buy SRI	0.37	0.48	0.00	1.00	652
do not (plan to) hold SRI	0.37	0.48	0.00	1.00	652
native speaker	0.77	0.42	0.00	1.00	652
born in Germany	0.76	0.43	0.00	1.00	652
age	29.22	8.08	18.00	72.00	652
female	0.29	0.45	0.00	1.00	652
income	3.25	2.00	1.00	7.00	652
highschool	0.34	0.47	0.00	1.00	652
apprenticeship	0.11	0.31	0.00	1.00	652
higher education	0.56	0.50	0.00	1.00	652
full-time	0.47	0.50	0.00	1.00	652
part-time	0.13	0.34	0.00	1.00	652
student	0.34	0.47	0.00	1.00	652
no employment	0.06	0.24	0.00	1.00	652

Notes: Table 19 presents summary statistics on all participants. Definitions of these variables can be found in Table 26 (appendix). The belief variables represent pre-treatment elicitations and are winsorized at the 1st and 99th percentiles.

3.3 Return Expectations, Asset Ownership, and the Influence on Sustainability Convictions

Our primary objective is to examine the consistency of investors' convictions of what sustainability should entail in the context of motivated beliefs. These beliefs are shaped by (i) return expectations and (ii) knowledge of the sustainability attributes of investor´s portfolio holdings within the experiment. By exogenously varying these factors, we investigate their causal impact on investors' sustainability convictions. Unless stated otherwise, all analyses were pre-registered before data collection at the American Economic Association's registry for randomized controlled trials under AEARCTR-0012566.

3.3.1 Prior Beliefs and Integrity of Randomization

Figure 8 displays the 12-month pre-treatment return point estimates for ETFs adhering to four different sustainability definitions (*alignment 2°C goal; exclusion of controversial firms; low emission intensity; no sustainability definition*). As shown in the graph the distribution is narrow, with few outliers. The average values are in a reasonable range between 6% and 9% (Table 19) compared to the average annual return of the most popular German stock market index (DAX) over the past 10 years with 7.13%.

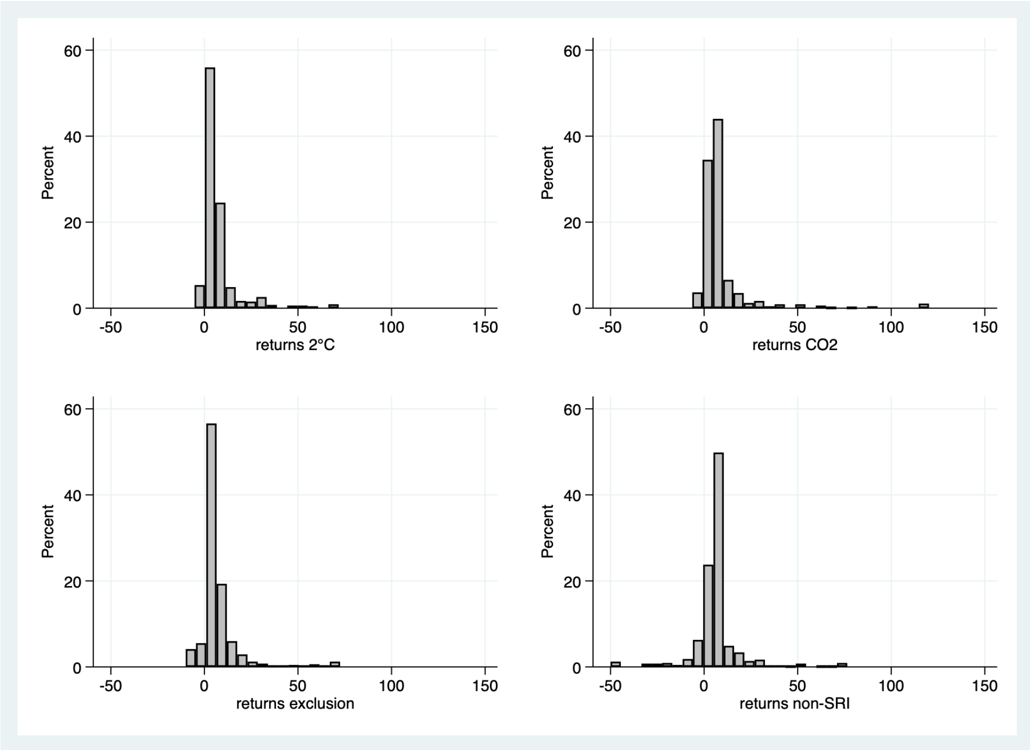


Figure 8: 12-month return estimates (pre-treatment)

Notes: This Figure presents the 12-month pre-treatment return point estimates for ETFs adhering to four different sustainability definitions. The top left graph displays the estimates for ETFs aligned with the 2°C goal definition. The top right graph shows the returns for ETFs with low emission intensity. The bottom left graph illustrates the returns for ETFs that exclude controversial firms. Finally, the bottom right graph represents ETFs that do not follow any of the sustainability definitions.

Overall our sample is well-balanced across the *return* as well as the *choice treatment*, considering a range of demographics and investment experience, as shown in the appendix Tables 27 and 28. Crucially, we observe no pre-treatment differences in return expectations between the treatments. However, there might be some imbalances in return uncertainty, holdings of sustainable stocks, education, and employment status. To address any concerns, we include a set of control variables in our main specifications.[35]

Result 1: Pre-treatment return beliefs largely mirror the past performance of the German DAX. We find no pre-treatment differences in return beliefs between treatments.

3.3.2 Updating of Return Expecations

A necessary condition for our subsequent analysis is that the *return treatment* influences return expectations of assets complying with the 2°C goal sustainability definition, while return expectations for the other sustainability definitions are largely unaffected. After the treatment, we ask participants to rate their agreement with three statements capturing return expectations over the next 12 months of ETFs following different sustainability definitions. Table 20 reports

[35]See the appendix for the pre-registered specifications without control variables. The results are broadly consistent.

OLS estimates of the effect of the *return treatment* on participants' posterior (change in) return expectations. We demonstrate three pieces of evidence that our treatment worked.

Table 20: The *return treatment* changes expected returns of 2°C goal aligned ETFs, while it leaves the return expectations of other sustainability indicators unaffected.

	(1) returns 2°C	(2) Δ returns 2°C	(3) returns CO2	(4) Δ returns CO2	(5) returns exclusion	(6) Δ returns exclusion
positive	0.627***	0.359***	-0.127	-0.00823	-0.0356	-0.00718
	(0.0739)	(0.0766)	(0.0783)	(0.0776)	(0.0779)	(0.0779)
Constant	-0.335***	-0.194***	0.0605	0.0156	0.00715	-0.00569
	(0.0516)	(0.0530)	(0.0538)	(0.0510)	(0.0563)	(0.0570)
N	652	652	652	652	652	652
adj. R^2	0.0982	0.0312	0.00248	-0.00152	-0.00122	-0.00153

Notes: The *return treatment* worked. Participants in the *positive treatment* group report higher posterior (stronger updating of) return expectations of 2°C aligned ETFs than participants in the *negative treatment* group (OLS regression). Return expectations (updating) of assets aligned with a low emission intensity or excluding controversial firms are unaffected. *returns 2°C* is the answer to the question *"ETFs aligned with the 2°C goal will generate higher returns over the next 12 months than ETFs not aligned with the 2°C goal."* on a scale from one (fully disagree) to five (fully agree) (post-treatment). *returns CO2* is the answer to the question *"ETFs with low emission intensity will generate higher returns over the next 12 months than ETFs with high emission intensity."* on a scale from one (fully disagree) to five (fully agree) (post-treatment). *returns exclusion* is the answer to the question *"ETFs that exclude controversial companies will generate higher returns over the next 12 months than ETFs that do not exclude controversial companies."* on a scale from one (fully disagree) to five (fully agree) (post-treatment). Δ *returns 2°C*, Δ *returns CO2*, and Δ *returns exclusion* indicate the difference between posterior and prior return expectation as explained in section 3.3.1. All dependent variables are z-scored. The variable *positive* is one if the subject is part of the positive returns treatment group (learns about a positive correlation between *alignment with the 2°C goal* and (expected) returns) and zero otherwise. Standard errors are displayed in parentheses. Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively.

First, Column (1) indicates that posterior return expectations differ between our treatment groups. The results show that agreement with the statement *"ETFs that are aligned with the 2°C target will generate higher returns over the next 12 months than ETFs that are not aligned with the 2°C target"* is 62.7% of a standard deviation higher in the *positive* than in the *negative treatment* group ($p < 0.01$). Second, Column (2) demonstrates that participants update their return beliefs due to our treatments. To mitigate concerns about experimenter demand effects, confusion (Haaland *et al.* 2023), and consistency bias (Falk and Zimmermann 2013), we follow Coibion *et al.* (2022) and use different elicitation techniques for prior (point estimates) and posterior (agreement with statement) return expectations (see Haaland *et al.* (2023) for a review). For the within-person comparison, we scale prior beliefs in the same way as posterior beliefs to assess updating. To do so, we follow our pre-analysis plan and calculate the average overperformance return expectations for each sustainability definition.[36] Δ *returns 2°C* is defined as the difference between posterior and prior return expectations. Column 2 shows that participants in the *positive treatment* increase their return beliefs by 35.9% of a standard deviation ($p < 0.01$)

[36]Let R_i be the expected return for sustainability definition i, and let $\bar{R}_{-i}$ be the average expected return of all other sustainability definitions and non-SRI. The term $\Delta R_i = R_i - \bar{R}_{-i}$ represents the average overperformance of sustainability definition i. We closely follow our pre-analysis plan and code the overperformance from 1 (overperformance of less than -5%), 2 (overperformance between -5% and -1%), 3 (overperformance between -1% and +1%), 4 (overperformance between +1% and +5%) to 5 (overperformance of more than +5%).

compared to the *negative treatment* group. Thus, we conclude that participants update their return beliefs due to our return treatment.

Third, Columns (3) and (5) indicate that posterior return expectations for low emission intensity ETFs or those excluding controversial companies are not significantly higher in the *positive* compared to the *negative treatment* group ($p = 0.11$, respectively $p = 0.65$).[37] Furthermore, the *positive treatment* did not lead to belief updates across these sustainability definitions (Columns (4) and (6); $p > 0.50$). Taken together, these findings suggest that the *positive* vs. the *negative treatment* group significantly increases return beliefs for ETFs aligned with the 2°C goal, while ETFs associated with other sustainability definitions remain largely unaffected. These results demonstrate the effectiveness of our experimental manipulation and are the foundation for our subsequent analysis.

Result 2: The *positive treatment* compared to the *negative treatment* significantly increases participants' return expectations over the next 12 months for ETFs aligned with the 2°C goal. Meanwhile, return expectations for all other ETFs remain mostly unaffected.

3.3.3 Belief Updating and the Consequences on the Choice of Sustainable Investments

Existing literature indicates that many investors value sustainability (Hartzmark and Sussman 2019; Bauer *et al.* 2021; Braun *et al.* 2024b), yet the drivers of sustainable investments are debated. Standard economic theory suggests individuals are incentivized by financial returns (Markowitz 1952), while recent literature shows that socially responsible investors are at least partly motivated by altruistic motives (Riedl and Smeets, 2017; Bauer *et al.*, 2021). We start our analysis by answering the question of whether investors' return expectations influence participants' choice of a sustainable investment among different sustainable investment options.[38] After allocating 50 euros between a non-sustainable ETF and a sustainable ETF, participants in the *active treatment group* actively choose up to three sustainability definitions that must be met by their sustainable investment. Participants are then randomly assigned an ETF that aligns with these criteria. Table 21 presents the results of OLS estimates, demonstrating an increased preference for investments aligned with the 2°C goal among participants in the *positive treatment* compared to those in the *negative treatment.*

The coefficient in Column (1) shows a 69 percentage points increase in the likelihood of choosing an investment aligned with the 2°C goal definition in the *positive treatment* ($p < 0.01$), which is 6.5 times larger than in the *negative treatment* group. This pattern persists when accounting for the share of sustainable investments (Column (2), $p < 0.01$) and demographic

[37]If anything, the *positive treatment* is associated with a decrease in return expectations for ETFs with low emission intensity (Specification (3), $p = 0.11$), which even strengthens the effect of our treatment.

[38]See Braun *et al.* (2024b) for a discussion on the utility of pecuniary and non-pecuniary motives.

Table 21: The choice of ETFs aligned with the 2°C goal definition is higher in the *positive* than in the *negative treatment* group.

	(1) 2°C choice	(2) 2°C choice	(3) 2°C choice
positive	0.693***	0.675***	0.675***
	(0.0400)	(0.0410)	(0.0437)
Constant	0.127***	-0.00898	-0.0473
	(0.0265)	(0.0387)	(0.140)
N	324	324	324
controls	no	yes	yes
split	active	active	active
adj. R^2	0.479	0.502	0.492

Notes: The OLS specification shows that participants in the *positive treatment* are more likely to choose ETFs aligned with the 2°C goal definition than those in the *negative treatment*. The dependent variable *2°C choice* is one if the subject actively selects *alignment with the 2°C goal* as sustainability definition in the interactive interface and therefore owns an assets in line with this definition and zero otherwise. The variable *positive* takes the value of one if the subject is in the *positive treatment group* and zero otherwise. Column (1) includes no controls, Column (2) adds the amount of sustainable investment, and Column (3) further includes a set of demographic controls. This analysis focuses exclusively on participants in the *active treatment*, as only those have the choice to select a sustainability definition. Significance levels are denoted as follows: *, **, and *** for the 10, 5, and 1 percent levels, respectively.

variables (Column (3), $p < 0.01$).[39] Taken together with the result that post-treatment return expectations for ETFs aligned with the 2°C goal are higher in the *positive* than in the *negative treatment* group, we infer that changes in return expectations influence sustainable investment decisions.

Result 3: Changes in return expectations significantly impact sustainable investment choices, with participants favoring options projected to yield higher returns.

3.3.4 Belief Updating and the Consequences on the Perception of Sustainability Definitions

Upon establishing the influence of returns on sustainable investments, one could argue that such behavior is rational. In the model proposed by Levitt and List (2007), investors' utility is derived from wealth and moral components, with investors seeking to maximize their utility through risk-adjusted returns (Markowitz 1952) and satisfaction from non-pecuniary motives (Andreoni 1989, 1990). To disentangle potentially rational return considerations in choosing sustainable investments from motivated beliefs that change the assessment of the sustainability attributes of an asset itself, our subsequent analysis focuses on the assessment of what sustainability should entail. More specifically, what do participants determine as the most suitable definition for classifying sustainable investments? Apart from the academic relevance of motivated beliefs for decision-making, it might be especially relevant for policymakers to consider these beliefs

[39]We include the amount of sustainable investment and a set of demographics to control for pre-treatment differences. The specifications (2) and (3) were not pre-registered.

in regulatory frameworks. Gino *et al.* (2016) argue that motivated beliefs strive in ambiguous surroundings. The different sustainability definitions according to MiFID II potentially are such an environment. We hypothesize that risk and return expectations foster motivated beliefs on the optimal sustainability definition and therefore overshadow genuine sustainability preferences.

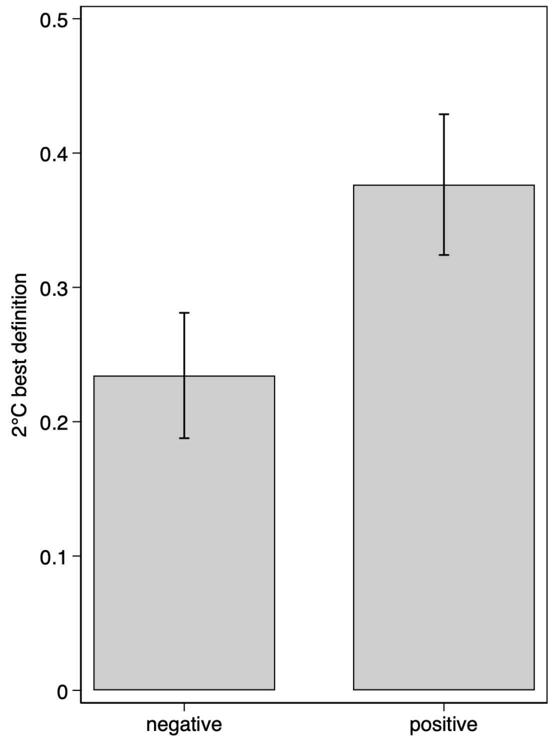


Figure 9: Impact of altered return expectations on sustainability assessments (pooled across *choice treatment*)
Notes: This Figure illustrates the share of participants choosing *alignment with the 2°C goal* as best sustainability indicator, differentiated by *positive* and *negative treatment* group. All results are pooled over *active* and *passive treatment* group. The error bars represent 95% confidence intervals.

To evaluate this question, we analyze the consistency of these sustainability evaluations in the *positive* versus the *negative treatment* group. The classification of sustainable investments should exclusively be guided by reliable data and measurements, compatibility with science-based targets, forward-looking perspectives, and additionality (Popescu *et al.*, 2021), while return expectations should not influence this decision. Figure 9 presents the results of this evaluation differentiated by *negative* and *positive treatment* group for the full sample. The proportion of participants who identify alignment with the 2°C goal as the most appropriate definition for classifying sustainable investments is 23.4 percentage points in the *negative treatment* compared to 37.6 percentage points in the *positive treatment* group ($p < 0.01$), a substantial increase of 60.7%.

Table 22: Participants in the *positive treatment* evaluate the *alignment with the 2°C goal* definition more favorably than those in the *negative treatment.*

	(1) 2°C best definition	(2) 2°C assessment	(3) 2°C best definition (arguments)	(4) Δ 2°C positive arguments
positive	0.142***	0.253***	0.106***	0.0643*
	(0.0367)	(0.0793)	(0.0385)	(0.0377)
Constant	0.206*	0.232	0.363***	0.318**
	(0.118)	(0.254)	(0.133)	(0.133)
N	652	652	500	630
controls	yes	yes	yes	yes
adj. R^2	0.0185	0.0251	0.0147	0.0332

Notes: The OLS specification shows that participants in the *positive treatment* assess the *alignment with the 2°C goal* definition more favorably than those in the *negative treatment* group. *2°C best criterion* is one if participants answer with *alignment with the 2°C goal* to the question *"Please select the criterion that you believe is the most important for evaluating sustainable investments."* and zero otherwise. *2°C assessment* reflects participants' evaluation of *alignment with the 2°C goal* relative to *exclusion of controversial companies* and *low emission intensity.* Higher values indicate a stronger conviction that the *alignment with the 2°C goal* is more convincing. Each measure's persuasiveness is elicited on a Likert scale from one to five, before the mean of the other two definitions is subtracted from the *alignment with the 2°C goal* score. The variable *2°C best definition (arguments)* is set to one if participants provide arguments favoring the 2°C goal sustainability definition as best to classify sustainable investments, and zero if they argue in favor of one of the other sustainability definitions. The variable Δ *2°C positive arguments* denotes the net count of arguments supporting the aligned with the 2°C goal definition. The latter three variables are z-scored. The explaining variable *positive* is one if the subject is part of the *positive treatment* group and zero otherwise. Standard errors are displayed in parentheses. Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

Table 22 displays OLS specifications including controls and corroborates the earlier result.[40] Specification (1) demonstrates that the share of participants who identify alignment with the 2°C goal as the most appropriate definition for classifying sustainable investments is 14.2 percentage points higher in the *positive treatment* than in the *negative treatment* group ($p < 0.01$). Furthermore, we leverage participants' evaluation of the sustainability definitions on a likert scale from 1 to 5 to construct the variable *2°C goal assessment*, which indicates if participants evaluate *alignment with the 2°C goal* more favorably than other sustainability definitions. The result demonstrates that participants in the *positive treatment* evaluate *2°C goal assessment* 25.3% of a standard deviation more favorably than participants in the *negative treatment* group ($p < 0.01$). On the evaluation page for each definition of sustainability, we present participants with both supporting and opposing arguments. They are then asked to explain their assessment of each definition. In addition, on the page evaluating the most appropriate definition for classifying sustainable investments, we ask participants to explain their reasons for choosing a particular sustainability definition. Most participants mention one or more of the arguments provided in the text, while only a minor fraction of participants (7.2%) cite return considerations as a factor in one of the sustainability assessments. Even if we follow the most conservative approach and exclude participants that cite returns in the respective sustainability evaluation, our results remain consistent (Table 29, appendix, Columns (2) and (4); $p < 0.01$). Although

[40]We include control variables to address concerns regarding pre-treatment differences. You find the pre-registered regression without controls in the appendix in Table 22, Column (1). The result is the same ($p < 0.01$).

most participants do not argue return expectations as a driver for the classification of sustainable investments, they both intentionally and unintentionally do include them in their assessment.

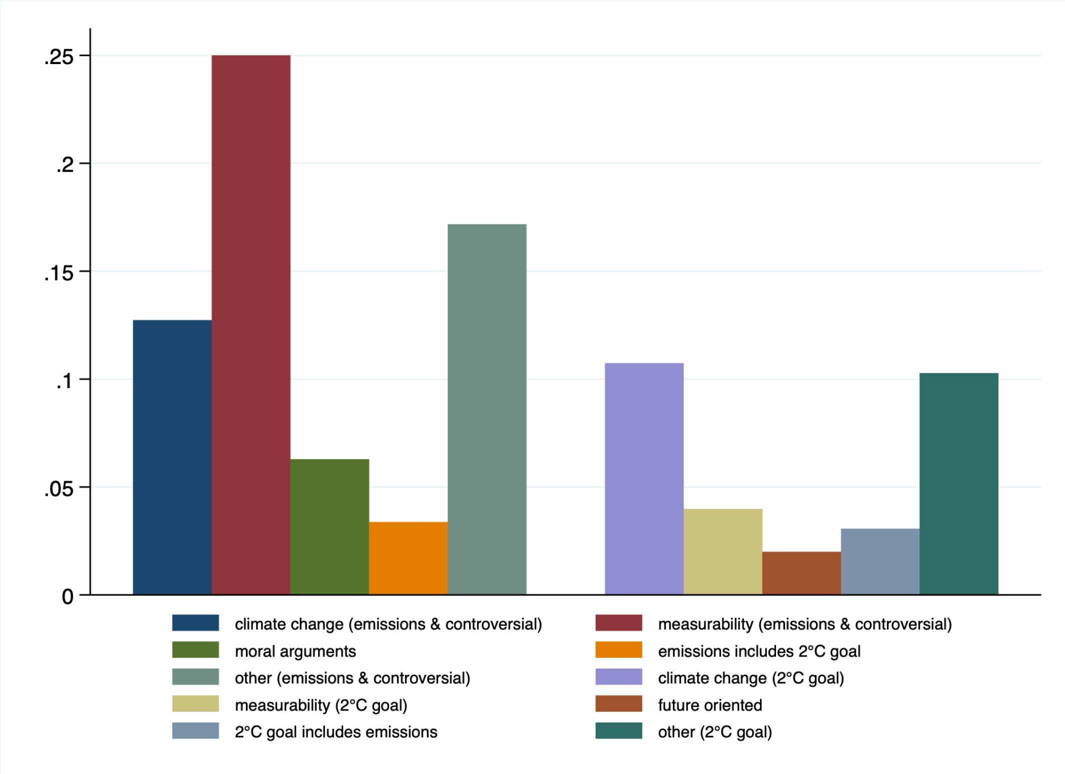


Figure 10: Participants reasoning for their endorsement of a certain sustainability definition
Notes: This Figure presents participants' categorized answers in the open text field explaining their decision why they vote for *exclusion of controversial companies*, *low emission intensity* or *alignment with the 2°C goal* as best sustainability definition. On the x-axis, you see the ten different categories. The first five bars represent arguments in favor of the *exclusion of controversial companies* or *low emission intensity* definition and the latter five bars represent arguments in favor of the *alignment with the 2°C goal* sustainability definition. The y-axis represents the share of participants that mention the respective arguments.

To dig deeper into participants' reasoning, we organize all open-text field answers into distinct categories following a preliminary analysis of the content.[41] Figure 10 illustrates the distribution of arguments for choosing a particular sustainability definition as the most suitable for classifying sustainable investments. Bars one to five represent arguments favoring *low emission intensity* and the *exclusion of controversial companies*, whereas bars six to ten depict those in support of aligning with the 2°C goal as the preferred future sustainability definition. Participants predominantly favor the *exclusion of controversial firms* or those with *low emission intensity*, citing their role in climate change mitigation (13%), the high transparency and good measurability of these criteria (25%), and moral reasons (6%). Additionally, 3% of participants argue that the *low emission intensity* definition also contributes to achieving the 2°C goal, and 17% provide other arguments supporting the mentioned sustainability definitions. Support for *alignment with the 2°C goal* is justified by its role in mitigating climate change (11%), measurability (4%), future orientation (2%), and the belief that achieving the 2°C goal implies lower emissions (3%).

[41]The following analysis is purely exploratory.

Moreover, 10% of participants mention other reasons supporting the 2°C goal definition. We construct an indicator set to one if an individual supports, and zero if an individual opposes, *alignment with the 2°C goal* as the best sustainability definition. For this analysis, we exclude all participants from whom we cannot determine whether their arguments are in favor of or against the 2°C goal sustainability definition. Specification (3) in Table 22 shows that participants in the *positive treatment* are 12 percentage points more likely to support the *alignment with the 2°C goal* definition than those in the *negative treatment* group ($p < 0.01$), representing a 68% increase. This result highlights that participants utilize the provided arguments. Despite receiving identical arguments for and against each sustainability definition, participants apply these arguments differently to justify their choices.

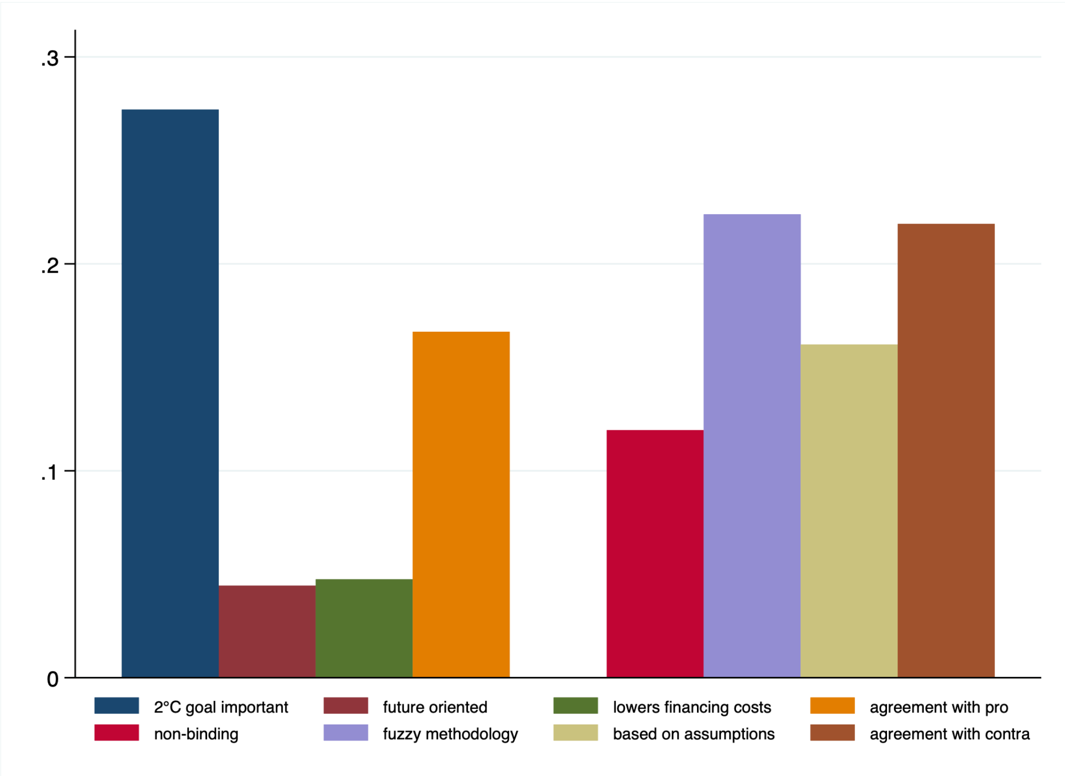


Figure 11: Participants reasoning in favor and against the *alignment with the 2°C goal* sustainability definition.
Notes: This Figure presents participants' categorized answers in the open text field explaining their decision in the evaluation of the 2°C goal definition. On the x-axis, you see the eight different categories. The first four bars represent arguments in favor and the latter four bars represent arguments against the proposed sustainability definition. The y-axis represents the share of participants that use the respective argument.

After exploring the reasons why participants chose *alignment with the 2°C goal* as the best sustainability definition, we further investigate their rationale for assessing whether *alignment with the 2°C goal* constitutes a suitable sustainability definition in itself. This analysis sheds light on the multifaceted motivations behind participants' sustainability evaluations. Figure 11 presents the distribution of arguments. Each response was categorized by a research assistant without knowledge of the hypotheses into one to three groups, reflecting the six arguments introduced during the experiment. Responses that went beyond the provided arguments were classified as either 'other pro' or 'other contra' arguments to ensure a comprehensive capture of participant perspectives. The majority of participants argued in favor of the *alignment with the 2°C goal* definition by highlighting its critical role in mitigating climate change risks (27%), its forward-looking approach (4%), and its ability to lower financing costs for compliant firms (5%). Additionally, 17% of the justifications highlight other motives in favor of the *alignment with the 2°C goal* definition, such as "sustainability is crucial," "it is the right thing to do," or "although the definition is not perfect, it is still better than doing nothing". Criticisms were primarily centered around nonbinding commitments (12%), methodological uncertainties (17%), and speculative assumptions (16%). Furthermore, 22% of the concerns mention other reasons against the *alignment with the 2°C goal* definition, like "global warming is unstoppable anyway," "it would be more effective to reduce emissions directly rather than in the future", or "limiting global warming to 2°C is insufficient".

In the last step of this analysis, we conduct a quantitative assessment of participants' qualitative responses, using an indicator variable, assigned a value of one when a participant provided more arguments in favor of than against the *alignment with the 2°C goal* definition and zero otherwise. We exclude all participants from whom no interpretable arguments are provided. The results are consistent with our previous findings. Specification (4) in Table 22 shows a 6.4 percentage point increase in the likelihood of participants advocating in favor of the *alignment with the 2°C goal* definition (Column (1), $p = 0.09$), representing a 21% increase compared to the *negative treatment* group.

For further evidence that the effect is a result of changes in return expectations, we correlate post-treatment return expectations and the four measures for the evaluation of the 2°C goal sustainability definition. Table 30 in the appendix demonstrates that the higher the ex-post return expectations of the 2°C goal aligned ETFs, the more likely it is that the 2°C goal definition is assessed as the most promising one to classify sustainable investments (the higher the rating of the 2°C goal definition compared to other sustainability definitions) (all specifications $p < 0.01$). Taken together, participant´s return expectations influence their sustainability convictions.

Result 4: Shifts in return expectations significantly alter the evaluation of the *alignment with the 2°C goal* sustainability definition. Despite all participants being presented with identical arguments, they selectively use and interpret these arguments in a self-serving manner to justify decisions that align with their motivated beliefs.

3.3.5 Return Expectations, Existing Sustainable Investment Decisions, and Sustainability Convictions

Previous to the experiment, we identified two potential drivers of motivated beliefs: (i) return expectations and (ii) sustainability attributes of existing portfolio holdings within the experiment. The latter may amplify motivated beliefs induced by return expectations because participants possess *skin in the game* (Taleb, 2018). Having demonstrated the effect of return expectations in the previous section, we now examine the combined effect of return expectations and existing portfolio holdings in a heterogeneity analysis. We hypothesize that the effects of motivated beliefs induced by the *positive treatment* are stronger in the *active treatment* than in the *passive treatment* group.

Table 23: The more favorable assessment of the *alignment with the 2°C goal* sustainability definition in the *positive treatment* group is driven by participants of the *active treatment* group.

	(1) 2°C best definition	(2) 2°C assessment	(3) 2°C best definition (arguments)	(4) Δ 2°C positive arguments
positive	0.0433	0.114	0.00134	0.0198
	(0.0515)	(0.113)	(0.0524)	(0.0526)
active	-0.0911*	-0.169	-0.0695	-0.0563
	(0.0484)	(0.112)	(0.0529)	(0.0492)
positive * active	0.201***	0.280*	0.128*	0.179**
	(0.0721)	(0.157)	(0.0752)	(0.0770)
Constant	0.256**	0.325	0.359***	0.390***
	(0.122)	(0.257)	(0.139)	(0.138)
N	652	652	630	500
controls	yes	yes	yes	yes
adj. R^2	0.0275	0.0270	0.0346	0.0231

Notes: The OLS regression shows the change in the assessment of the 2°C goal sustainability definition between the *positive treatment* and the *negative treatment* group is stronger in the *active treatment* compared to the *passive treatment* group. *2°C best criterion* is one if participants answer with *alignment with the 2°C goal* to the question *"Please select the criterion that you believe is the most important for evaluating sustainable investments."* and zero otherwise. *2°C assessment* reflects participants' evaluation of *alignment with the 2°C goal* relative to *exclusion of controversial companies* and *low emission intensity.* Higher values indicate a stronger conviction that the *alignment with the 2°C goal* is more convincing. Each measure's persuasiveness is elicited on a Likert scale from one to five, before the mean of the other two definitions is subtracted from the *alignment with the 2°C goal* score. The variable *2°C best definition (arguments)* is set to one if participants provide arguments favoring the 2°C goal sustainability definition as best to classify sustainable investments, and zero if they argue in favor of one of the other sustainability definitions. The variable Δ *2°C positive arguments* denotes the net count of arguments supporting the aligned with the 2°C goal definition. The latter three variables are z-scored. The variable *positive* is one if the subject learns about a positive correlation between *alignment with the 2°C goal* and (expected) returns) and zero otherwise. *active* is an indicator variable that is one if a subject is part of the group that can actively choose the sustainability criterion before the random selection of funds and zero otherwise. Standard errors are displayed in parentheses. Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

To explore this, we investigate how the *positive treatment* interacts with the *active treatment* group in influencing the selection of the optimal sustainability definition. A significant interaction effect would suggest that investors informed about their investments' sustainability attributes (*active treatment* group) respond more strongly to the *positive treatment* than those unaware of their portfolio's attributes (*passive treatment*). Table 23 presents the results of an OLS regression. The interaction term shows a significant positive effect for determining the 2°C goal as the most suitable sustainability definition (Column (1), $p < 0.01$) indicating that motivated beliefs are about five times stronger in the *active* than in the *passive treatment group*. On the contrary, the coefficient for the *positive treatment* is not statistically significant (Column (1), $p = 0.40$). This indicates no significant effect of the *positive treatment* on the *alignment with the 2°C goal* definition in the *passive treatment group*. Together these results suggest that the participants are mostly influenced by motivated beliefs induced through the interplay of knowledge on the sustainability attributes of the participant´s ETF and changed return expectations. Furthermore, the results remain statistically and economically significant if we use the assessment of the 2°C goal compared to the other two sustainability definitions or our two variables deducted by the qualitative answer of the participants as dependent variable (Columns (2) - (4); p-values between 0.01 to 0.09).[42]

A competing hypothesis might be that the effects observed are not primarily driven by motivated beliefs linked to the interplay of return expectations and existing investments but by enhanced learning about the sustainability-return correlation among participants in the *active treatment group*. This enhanced learning could arise from the participant´s opportunity to engage with an interactive interface displaying ETFs, their alignment with different sustainability definitions, and expected returns for a second time. Importantly, participants in the *active treatment group* are not given additional information beyond what is initially presented; they interact with the same interface that displays the ETFs and their sustainability-return correlations as on the previous page. Similarly, those in the *passive treatment group* are also revisited with the full list of ETFs and their respective sustainability-return correlations. The key difference is the interaction with the study's tool: participants in the *active treatment group* express their sustainability preferences by selecting their desired sustainability definitions through the interactive interface. Subsequently, they are matched with an ETF that aligns with these stated preferences. Conversely, participants in the *passive treatment group* are not allowed to indicate their sustainability preferences and therefore assigned a random sustainable ETF. Hence, these participants are not aware of the sustainability attributes of their investment.

To address potential concerns about learning effects within the *active treatment group*, we conduct various robustness tests using OLS regressions incorporating controls for prior and posterior return expectations, the time spent on the learning and decision pages, the number of clicks on these pages, as well as demographic variables (see Table 24).[43] First, we consider posterior return expectations to determine whether the interaction effects stem from learning-

[42]Table 31 shows the pre-registered specifications without controls (Specifications (1) and (3)) as well as specifications, where we exclude all individuals mentioning return expectations as the motivation behind their choice (assessment) of sustainability definitions. The results are largely the same.

[43]This robustness analysis was not pre-registered. We show it to reinforce confidence in the proposed mechanism.

Table 24: The higher likelihood of endorsing the *alignment with the 2°C goal* sustainability definition in the *positive treatment* group is driven by participants of the *active treatment* (controlling for posterior returns, time and # of clicks)

	(1) 2°C best definition	(2) 2°C best definition	(3) 2°C best definition	(4) 2°C best definition	(5) 2°C best definition
returns 2°C	0.125***	0.126***	0.126***	0.125***	0.125***
	(0.0175)	(0.0175)	(0.0175)	(0.0176)	(0.0177)
positive	-0.0110	-0.0155	-0.0166	-0.0165	-0.0154
	(0.0503)	(0.0506)	(0.0505)	(0.0505)	(0.0506)
active	-0.0293	-0.0498	-0.0474	-0.0685	-0.0711
	(0.0468)	(0.0518)	(0.0525)	(0.0594)	(0.0588)
positive * active	0.136*	0.143**	0.147**	0.144**	0.140**
	(0.0699)	(0.0700)	(0.0696)	(0.0695)	(0.0700)
time ETF selection		0.000593	-0.00162	-0.00173	
		(0.00101)	(0.00353)	(0.00356)	
time ETF selection^2			0.0000250	0.0000225	
			(0.0000377)	(0.0000380)	
# clicks learning				-0.00118	-0.00119
				(0.00220)	(0.00242)
# clicks decision				0.00458	0.00532
				(0.00667)	(0.00565)
delta time					0.0000399
					(0.000342)
Constant	-0.141	-0.136	-0.104	-0.0760	-0.0934
	(0.144)	(0.146)	(0.153)	(0.156)	(0.151)
N	652	646	646	646	646
controls	yes	yes	yes	yes	yes
adj. R^2	0.101	0.102	0.102	0.0997	0.101

Notes: The OLS regression shows the persistent interaction effect of the *positive* and *active treatment* group on the endorsement of the *alignment with the 2°C goal* controlling for posterior returns, time, and number of clicks. *2°C best definition* is one if participants endorese *alignment with the 2°C goal* to the question as best sustainability definition and zero otherwise. The variable *positive* is one if the subject learns about a positive correlation between *alignment with the 2°C goal* and (expected) returns and zero otherwise. *active* is an indicator variable that is one if a subject is part of the group that can actively choose the sustainability definition before the random selection of funds and zero otherwise. *time ETF selection* measures the duration spent on the fund selection page, and this variable is winsorized at the 5% and 95% levels. The variables *# clicks learning* and *# clicks decision* represent the number of clicks on different sustainability criteria in the learning (fund selection page) and are winsorized at the 90% level. *delta time* measures the time difference between the learning and decision screens and is winsorized at the 10% and 90% levels. All time variables are winsorized on the 10% and 90% levels () In all specifications, we adjust for demographics and pre-treatment return expectations. Standard errors are displayed in parentheses. Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

induced changes in return expectations or from the choice of 2°C goal-aligned assets. Although the interaction effect's magnitude decreases compared to our standard specification, a statistically significant interaction remains (Column (1), $p < 0.052$), indicating the importance of the sustainability attributes of the existing portfolio holdings within the experiment. Second, to address concerns regarding the time spent by participants in the *active treatment group* on the decision screen, we include controls for both linear and quadratic time effects and show the robustness of our results (Columns (2) and (3), $p = 0.041$ and $p = 0.035$, respectively).[44] Third, including the number of clicks on sustainability definitions on both the learning and decision pages does not change the significance of the interaction effect (Column (4), $p = 0.038$), demonstrating the robustness of the effect to variations in engagement with the sustainability-return information. Fourth, the results persist when we adjust for the time difference between the decision and learning screens (Column (5), $p = 0.045$), reinforcing the persistent influence of participants´ existing portfolio holdings on the selection of sustainability definitions. Table 32 (appendix) qualitatively supports the same conclusion for the evaluation of the *alignment with the 2°C goal* compared to the other sustainability definitions; the interaction effect remains positive although it becomes statistically insignificant. In summary, our findings consistently show that knowledge about the sustainability attributes of existing portfolio holdings significantly affects participants' preferences for incorporating the 2°C goal into future sustainability definitions. This effect cannot be attributed solely to enhanced learning.

So far, we have demonstrated that altered return expectations and knowledge about the sustainability attributes of portfolio holdings influence the evaluation of the alignment with the 2°C goal definition to classify sustainable investments. To further investigate the mechanisms, we examine participants possessing 2°C goal-aligned ETFs and compare them with those (i) not owning these ETFs and (ii) those not knowing the attributes of their sustainable assets. If the observed effects are driven by participants actually owning 2°C goal-aligned ETFs, then, conditional upon being in the *positive (negative) treatment* group, participants choosing assets aligned with the 2°C goal are expected to show a preference for this definition. Table 25 supports this, indicating a pronounced preference for the alignment with the 2°C goal definition among such asset owners.[45] In particular, participants in the *positive treatment* owning 2°C goal-aligned assets are 74% more likely to support this goal as the optimal sustainability definition, compared to those not owning 2°C goal-aligned assets (see Column (1), $p = 0.03$). Moreover, when comparing 2°C-goal aligned asset owners to those unaware of their assets' sustainability attributes, the former are 47% more inclined to adopt the 2°C goal as the most fitting sustainability definition (see Column (2), $p < 0.01$). In contrast, among participants in the *negative treatment* group, no significant preference difference is observed, irrespective of their asset ownership (see Column (3), $p = 0.55$; Column (4), $p = 0.84$).

Table 33 (appendix) reaffirms this finding for an alternative specification with the *2°C assessment* as the dependent variable with the difference that the effect also holds in the *negative treatment*. Note, there exist no pre-treatment differences in prior return expectations between participants who own 2°C goal-aligned assets and those who do not (unreported tests; $p = 0.98$)

[44] For 6 individuals we do not have information on the time spent on the different pages. Our results are robust to the imputation of the median time and the removal of these individuals from the sample. Furthermore, in unreported specifications, we also account for the time spent on the learning page and for deciles of the time spent on the decision page. The results remain consistent.

[45] We pre-registered to analyze heterogeneity concerning the ETF holdings, but the exact analysis is exploratory.

Table 25: The selection of 2°C aligned ETFs increases the likelihood of endorsing this definition as best suitable for classifying sustainable investments.

	(1) 2°C best definition	(2) 2°C best definition	(3) 2°C best definition	(4) 2°C best definition
owns 2°C goal asset	0.204**	0.151***	0.0616	-0.0216
	(0.0919)	(0.0562)	(0.103)	(0.104)
Constant	0.267***	0.319***	0.188***	0.272***
	(0.0812)	(0.0363)	(0.0335)	(0.0351)
N	166	302	158	182
controls	no	no	no	no
split	positive /active	positive	negative /active	negative
adj. R^2	0.0191	0.0206	-0.00373	-0.00532

Notes: The OLS regression indicates that ownership of ETFs *Aligned with the 2°C goal* sustainability definition increases the likelihood of endorsing this definition as best suitable for classifying sustainable investments. This effect is driven by participants in the *positive treatment*. *2°C best criterion* is one if participants answer with *alignment with the 2°C goal* to the question *"Please select the criterion that you believe is the most important for evaluating sustainable investments."* and zero otherwise. The variable *own 2°C goal asset* is an indicator that is one if a participant is aware that she owns an asset aligned with the 2°C goal definition and zero otherwise. Columns (1) and (2) exclusively evaluate participants of the *positive treatment*, while Columns (3) and (4) focus on the negative treatment. Furthermore, Columns (1) and (3) evaluate participants that are aware of the attributes of their sustainable assets (*active treatment* group), while Columns (2) and (4) compare those owning a 2°C goal asset with those not knowing the attributes of their sustainable assets (*passive treatment* group). Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively.

or are not aware of their portfolio holdings (unreported tests, $p = 0.17$). Furthermore, our results stay the same if we include pre-treatment return expectations in all specifications (results available upon request), which shows that the results are not a consequence of pre-treatment differences. Overall, the findings highlight the crucial role of ownership of assets in shaping preferences for sustainability definitions, especially among participants with positive return expectations in respect of 2°C goal-aligned assets.

Result 5: The evaluation of the 2°C goal as a sustainability definition is influenced by changes in return expectations. This effect is particularly strong for participants owning 2°C goal-aligned assets.

3.4 Discussion and Conclusion

In our investigation, we meticulously design an online experiment to explore how investors' sustainability convictions are influenced by motivated beliefs. Respondents significantly update their beliefs regarding the performance of different ETFs due to information on the expected sustainability-return correlation. This allows us to analyze the effect of changes in return expectations on altering sustainability convictions. Furthermore, by altering participants' capacity to know the sustainability attributes of their portfolio holdings within the experiment, we investigate the combined effect of existing portfolio holdings and altered return expectations on participants' perceptions of what constitutes sustainability.

Our findings provide causal evidence that investors' sustainability convictions are shaped by motivated beliefs, stemming from the interplay between the sustainability attributes of existing portfolio holdings and return expectations. We demonstrate through open-text responses that despite receiving identical arguments for and against, participants interpret and apply these arguments differently to justify their choices. This highlights the need for policymakers to consider motivated beliefs in regulatory frameworks. We suggest that MiFID II regulations adopt a stricter procedure for eliciting sustainability preferences to mitigate this channel of self-serving behavior. Current regulations offer excessive flexibility in classifying sustainable investments, potentially allowing risk and return considerations, as well as the attributes of existing portfolio holdings, to overshadow genuine sustainability preferences. As our experiment shows, investors engage in self-serving behavior and rationalize the investments most aligned with their personal (financial) interests as sustainable. Therefore, we advocate for narrowing down sustainability definitions to the minimum, ideally to a single, universal definition based on clear indicators because motivated beliefs thrive in environments of ambiguity (Gino *et al.*, 2016). The tendency of individuals to choose sustainability definitions that match their return expectations and the sustainability attributes of their existing portfolio holdings is intensified by the European Commission's policy, which allows clients full freedom in determining what sustainability means to them. Furthermore, it must be ensured that advisors cannot influence their clients in the sustainability assessment. As shown by Laudi *et al.* (2023), advisors exploit the sustainability preferences of their clients by charging especially low financially literate investors higher fees. Therefore, it could well be that advisors use the flexibility offered by different sustainability definitions to sell products with high fees, arguing that they align with the previously manipulated sustainability preferences of their clients.

While our study provides valuable insights, it is bound by the constraints of its experimental design and participant demographics. Future studies could extend our analysis to real-world investment behaviors, examining the enduring impact of motivated beliefs on actual portfolio decisions. Furthermore, future research could explore if motivated beliefs can be attributed to a previous sustainable asset choice, the ownership of an asset itself, or a combination of both. Moreover, given that politicians may not favor a universal sustainability definition with limited factors, future research might also explore the effects of additional sustainability definitions and delve deeper into how a future sustainability definition should be crafted. Our study documents the effects of motivated beliefs on the process of shaping sustainability preferences. We leave the question of which factors need to be included in a holistic sustainability definition to future research and policymakers.

3.5 Appendix

Table 26: Variable definitions

Variable	Definition
2°C best definition	Indicator variable that is one if participants answer with *alignment with the 2°C goal* to the question *"Please select the criterion that you believe is the most important for evaluating sustainable investments."* and zero otherwise.
2°C assessment	Represents the difference between the seven-point likert scale evaluation of the *alignment with the 2°C goal* definition and the average rating of the other two definitions.
positive	Indicator variable that is one if a subject is part of the *positive treatment* group (learns about a positive correlation between *alignment with the 2°C goal* and (expected) returns) and zero otherwise.
active	Indicator variable that is one if a subject is part of the *active treatment group* that can actively choose the sustainability criterion before the random selection of funds and zero otherwise.
returns 2°C (prior)	Point estimate on the expected return over the next 12 months of an ETF aligned with the 2°C goal (prior treatment).
returns 2°C (posterior)	Response to the question *"ETFs aligned with the 2°C goal will generate higher returns over the next 12 months than ETFs not aligned with the 2°C goal."* on a scale from one (fully disagree) to five (fully agree) (post treatment).
Δ *returns 2°C*	Standardized difference between posterior and prior return expectations of 2°C goal ETFs.
returns exclusion (prior)	Point estimate on the expected return over the next 12 months of an ETF that excludes controversial companies (prior treatment).
returns exclusion (posterior)	answer to the question *"ETFs that exclude controversial companies will generate higher returns over the next 12 months than ETFs that do not exclude controversial companies."* on a scale from one (fully disagree) to five (fully agree) (post treatment).
Δ *returns exclusion*	Standardized difference between posterior and prior return expectations of ETFs that consider exclusion criteria.
returns CO2 (prior)	Point estimate on the expected return over the next 12 months of an ETF with a low emission intensity (prior treatment).
returns CO2 (posterior)	Answer to the question *"ETFs with low emission intensity will generate higher returns over the next 12 months than ETFs with high emission intensity."* on a scale from one (fully disagree) to five (fully agree) (post treatment).
Δ *returns CO2*	Standardized difference between posterior and prior return expectations of ETFs that consider exclusion criteria.
returns non-SRI	Point estimate on the expected return over the next 12 months of an ETF with no considerations of sustainability (prior treatment).
return uncertainty	Degree of confidence on the return answer on a scale from one (very insecure) to five (very secure).
investment experience	Experience in investments in stocks, bonds, and funds/ETFs (in years).

Table 26 (continued):

Variable	Definition
own SRI	Indicator variable that is one if the respondent currently holds sustainable investments and zero otherwise.
plans to buy SRI	Indicator variable that is one if the individual currently holds no sustainable assets but plans to buy some within the next three years and zero otherwise.
do not (plan to) hold SRI	Indicator variable that is one if the individual neither holds nor plans to hold SRI in the next three years.
native speaker	Indicator variable that is one if the respondent is a German native speaker and zero otherwise.
born in Germany	Indicator variable that is one if the individual was born in Germany and zero otherwise.
female	Indicator variable that is one if the respondent is female and zero otherwise.
higher education	Indicator variable that is one if the respondent has a higher education degree and zero otherwise.
high school	Indicator variable that is one if the respondent has a high school degree and zero otherwise.
apprenticeship	Indicator variable that is one if a respondent has an apprenticeship degree and zero otherwise.
age	Respondents age (in years)
income	Respondents self-reported net income: (1; below 1.000 euro) (2; 1.000 - 1,500 euro) (3; 1,501 - 2,000 euro) (4; 2,001 - 2,500 euro) (5; 2,501 - 3,000 euro) (6; 3,001 - 3,500 euro) (7; more than 3,500 euro)
full-time	Indicator variable that is one if a respondent works full-time and zero otherwise.
part-time	Indicator variable that is one if a respondent works part-time and zero otherwise.
student	Indicator variable that is one if a respondent is a student and zero otherwise.
no employment	Indicator variable that is one if a respondent is unemployed and zero otherwise.
time ETF selection	# of seconds on the ETF selection page.
# clicks learning	# number of clicks in the interactive tool on the learning page.
# clicks decision	# number of clicks in the interactive tool on the decision page.
delta time	Difference in the # of seconds between learning and decision page.
owns 2°C goal asset	Indicator variable that is one if a participant is aware that she owns an asset aligned with the 2°C goal definition and zero otherwise.
2°C best definition (arguments)	Indicator variable that is one if a participant provides more arguments in favor than against the *alignment with the 2°C goal* as best definition.
Δ *positive arguments*	Difference between arguments in favor and against the *alignment with the 2°C goal* definition.

Notes: Table 26 defines all relevant variables.

Table 27: No unexpected systematic differences in key observables between *positive* and *negative treatment* group.

	Positive	Negative	Diff.	Std. Error	P-value	Obs.
returns 2°C	7.81	7.34	-0.48	0.84	(0.569)	652
returns exclusion	6.98	7.67	0.68	0.91	(0.452)	652
returns CO2	9.03	9.95	0.92	1.26	(0.465)	652
returns non-SRI	5.80	6.59	0.80	1.05	(0.450)	652
return uncertainty	2.30	2.47	0.16**	0.08	(0.038)	652
investment experience	2.80	3.29	0.49	0.37	(0.181)	652
own SRI	0.22	0.31	0.09***	0.03	(0.007)	652
plans to buy SRI	0.35	0.39	0.04	0.04	(0.313)	652
do not (plan to) hold SRI	0.43	0.30	-0.13***	0.04	(0.001)	652
native speaker	0.76	0.78	0.02	0.03	(0.501)	652
born in Germany	0.75	0.76	0.01	0.03	(0.711)	652
age	29.55	28.88	-0.66	0.63	(0.297)	652
female	0.31	0.26	-0.04	0.04	(0.207)	652
income	3.34	3.17	-0.17	0.16	(0.264)	652
highschool	0.33	0.34	0.01	0.04	(0.865)	652
apprenticeship	0.11	0.10	-0.01	0.02	(0.826)	652
higher education	0.56	0.56	-0.00	0.04	(0.980)	652
full-time	0.47	0.46	-0.01	0.04	(0.730)	652
part-time	0.12	0.14	0.02	0.03	(0.446)	652
student	0.33	0.36	0.03	0.04	(0.405)	652
no employment	0.08	0.04	-0.04**	0.02	(0.048)	652

Notes: Table 27 verifies that the randomization into *positive* and *negative treatment* overall worked. To do so, mean values of all relevant variables are computed for each treatment group and compared between the two treatments by using t-tests. All belief variables represent pre-treatment elicitations and are winsorized at the 1st and 99th percentiles. Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively. For definitions of the variables, please consider Table 26.

Table 28: No unexpected systematic differences in key observables between *active* and *passive treatment* group.

	Active	Passive	Diff.	Std. Error	P-value	Obs.
returns 2°C	7.87	7.30	-0.57	0.84	(0.494)	652
returns exclusion	7.70	6.94	-0.77	0.91	(0.398)	652
returns CO2	9.61	9.35	-0.25	1.26	(0.841)	652
returns non-SRI	6.07	6.30	0.23	1.05	(0.824)	652
return uncertainty	2.35	2.41	0.06	0.08	(0.453)	652
investment experience	2.75	3.33	0.58	0.37	(0.117)	652
own SRI	0.25	0.27	0.02	0.03	(0.480)	652
plans to buy SRI	0.37	0.37	0.00	0.04	(0.967)	652
do not (plan to) hold SRI	0.38	0.35	-0.03	0.04	(0.492)	652
native speaker	0.78	0.76	-0.02	0.03	(0.511)	652
born in Germany	0.77	0.75	-0.02	0.03	(0.583)	652
age	28.96	29.48	0.51	0.63	(0.418)	652
female	0.30	0.27	-0.03	0.04	(0.335)	652
income	3.21	3.30	0.08	0.16	(0.597)	652
highschool	0.36	0.31	-0.05	0.04	(0.151)	652
apprenticeship	0.08	0.13	0.05**	0.02	(0.035)	652
higher education	0.56	0.56	0.00	0.04	(0.951)	652
full-time	0.47	0.46	-0.01	0.04	(0.762)	652
part-time	0.10	0.16	0.05*	0.03	(0.055)	652
student	0.35	0.33	-0.02	0.04	(0.544)	652
no employment	0.07	0.05	-0.02	0.02	(0.398)	652

Notes: Table 28 verifies that the randomization into *active* and *passive treatment* group worked. To do so, mean values of all relevant variables are computed for each treatment group and compared between the two treatments by using t-tests. All belief variables represent pre-treatment elicitations and are winsorized at the 1st and 99th percentiles. Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively. For definitions of the variables, please consider Table 26.

Table 29: Participants in the *positive treatment* are more likely to assess the *alignment with the 2°C goal* sustainability definition more favourably compared to the *negative treatment* group.

	(1) 2°C best definition	(2) 2°C best definition	(3) 2°C assessment	(4) 2°C assessment	(5) 2°C best definition (arguments)	(6) Δ 2°C positive arguments
positive	0.142***	0.142***	0.265***	0.244***	0.117***	0.0621
	(0.0357)	(0.0371)	(0.0783)	(0.0838)	(0.0375)	(0.0377)
Constant	0.234***	0.213*	-0.134**	0.187	0.173***	0.309***
	(0.0237)	(0.120)	(0.0565)	(0.255)	(0.0237)	(0.0264)
N	652	643	652	609	500	630
controls	no	yes	no	yes	no	no
split	no	no return	no	no return	no	no
adj. R^2	0.0222	0.0169	0.0159	0.0244	0.0174	0.00270

Notes: Notes: The OLS specification shows that participants in the *positive treatment* evaluate the *alignment with the 2°C goal* definition more favorably than those in the *negative treatment* group. *2°C best criterion* is one if participants answer with *alignment with the 2°C goal* to the question *"Please select the criterion that you believe is the most important for evaluating sustainable investments."* and zero otherwise. *2°C assessment* reflects participants' evaluation of *alignment with the 2°C goal* relative to *exclusion of controversial companies* and *low emission intensity*. Higher values indicate a stronger conviction that the *alignment with the 2°C goal* is more convincing. Each measure's persuasiveness is elicited on a Likert scale from one to five, before the mean of the other two definitions is subtracted from the *alignment with the 2°C goal* score. The variable *2°C best definition (arguments)* is set to one if participants provide arguments favoring the 2°C goal sustainability definition as best to classify sustainable investments, and zero if they argue in favor of one of the other sustainability definitions. The variable Δ *2°C positive arguments* denotes the net count of arguments supporting the aligned with the 2°C goal definition. The latter three variables are z-scored. The explaining variable *positive* is one if the subject is part of the *positive treatment* group and zero otherwise. Specifications (1), (2), (5), and (6) mirror Table 22 but do not include control variables. Specifications (2) and (4) exclude all individuals citing return expectations as a motivation for their evaluation of the sustainability definitions. Standard errors are displayed in parentheses. Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

Table 30: Participants' return expectations are positively correlated with their evaluation of the *alignment with the 2°C goal* sustainability definition.

	(1) 2°C best definition	(2) 2°C assessment	(3) 2°C best definition (arguments)	(4) Δ 2°C positive arguments
returns 2°C	0.151***	0.253***	0.0947***	0.122***
	(0.0179)	(0.0413)	(0.0202)	(0.0184)
Constant	0.207*	0.242	0.389***	0.296**
	(0.111)	(0.245)	(0.129)	(0.128)
N	652	652	500	630
controls	yes	yes	yes	yes
adj. R^2	0.0966	0.0698	0.0450	0.0916

Notes: This OLS regression shows that higher return expectations are positively correlated with a more favorable assessment of the 2°C goal sustainability definition. *2°C best criterion* is one if participants answer with *alignment with the 2°C goal* to the question *"Please select the criterion that you believe is the most important for evaluating sustainable investments."* and zero otherwise. *2°C assessment* reflects participants' evaluation of *alignment with the 2°C goal* relative to *exclusion of controversial companies* and *low emission intensity.* Higher values indicate a stronger conviction that the *alignment with the 2°C goal* is more convincing. Each measure's persuasiveness is elicited on a Likert scale from one to five, before the mean of the other two definitions is subtracted from the *alignment with the 2°C goal* score. The variable *2°C best definition (arguments)* is set to one if participants provide arguments favoring the 2°C goal sustainability definition as best to classify sustainable investments, and zero if they argue in favor of one of the other sustainability definitions. The variable Δ *2°C positive arguments* denotes the net count of arguments supporting the aligned with the 2°C goal definition. The latter three variables are z-scored. The variable *returns 2°C* is z-scored and indicates the post-treatment agreement with the statement *"ETFs aligned with the 2°C goal will generate higher returns over the next 12 months than ETFs not aligned with the 2°C goal."*. Standard errors are displayed in parentheses. Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

Table 31: The more favourable evaluation of the *alignment with the 2°C goal* sustainability definitions in the *positive treatment* group is driven by the participants of the *negative treatment* group.

	(1) 2°C best definition	(2) 2°C best definition	(3) 2°C assessment	(4) 2°C assessment
positive	0.0477	0.0450	0.117	0.0961
	(0.0505)	(0.0518)	(0.113)	(0.118)
active	-0.0754	-0.0871*	-0.157	-0.188
	(0.0473)	(0.0491)	(0.113)	(0.120)
positive * active	0.190***	0.198***	0.299*	0.307*
	(0.0710)	(0.0728)	(0.156)	(0.165)
Constant	0.272***	0.262**	-0.0560	0.289
	(0.0351)	(0.124)	(0.0814)	(0.261)
N	652	643	652	609
controls	no	yes	no	yes
split	no	no return	no	no return
adj. R^2	0.0304	0.0253	0.0184	0.0270

Notes: The OLS regression shows the change in the assessment of the 2°C goal sustainability definition between the *positive treatment* and the *negative treatment* group is stronger in the *active treatment* compared to the *passive treatment* group. *2°C best criterion* is one if participants answer with *alignment with the 2°C goal* to the question *"Please select the criterion that you believe is the most important for evaluating sustainable investments."* and zero otherwise. *2°C assessment* reflects participants' evaluation of *alignment with the 2°C goal* relative to *exclusion of controversial companies* and *low emission intensity.* Higher values indicate a stronger conviction that the *alignment with the 2°C goal* is more convincing. Each measure's persuasiveness is elicited on a Likert scale from one to five, before the mean of the other two definitions is subtracted from the *alignment with the 2°C goal* score. The variable is z-scored. The explaining variable *positive* is one if the subject is part of the *positive treatment* group and zero otherwise. Specifications (1) and (3) mirror Table 23 but do not include control variables (pre-registered specification). Specifications (2) and (4) exclude all individuals citing return expectations as a motivation for their evaluation of the sustainability definitions. Standard errors are displayed in parentheses. Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

Table 32: Although still positive, the interaction effect of *positive treatment* and *active treatment group* on the assessment of the 2°C goal sustainability definition is not statistically significant, controlling for posterior returns, time, and # of clicks.

	(1) 2°C assessment	(2) 2°C assessment	(3) 2°C assessment	(4) 2°C assessment	(5) 2°C assessment
returns 2°C	0.304***	0.305***	0.305***	0.297***	0.295***
	(0.0362)	(0.0363)	(0.0363)	(0.0365)	(0.0366)
positive	0.0143	0.0217	0.0234	0.0279	0.0240
	(0.104)	(0.105)	(0.105)	(0.105)	(0.105)
active	0.0474	0.0478	0.0441	-0.0176	-0.00317
	(0.103)	(0.111)	(0.111)	(0.124)	(0.123)
positive * active	0.0345	0.0419	0.0357	0.0205	0.0328
	(0.143)	(0.144)	(0.145)	(0.144)	(0.143)
time ETF selection		0.00000737	0.00341	0.00451	
		(0.00200)	(0.00721)	(0.00725)	
time ETF selection^2			-0.0000385	-0.0000572	
			(0.0000776)	(0.0000771)	
# clicks learning				-0.00981**	-0.0112**
				(0.00450)	(0.00485)
# clicks decision				0.0112	0.0103
				(0.0128)	(0.0111)
delta time					-0.000611
					(0.000649)
Constant	-0.828***	-0.810***	-0.859***	-0.703**	-0.696**
	(0.288)	(0.291)	(0.305)	(0.310)	(0.299)
N	652	646	646	646	646
controls	yes	yes	yes	yes	yes
adj. R^2	0.208	0.210	0.209	0.213	0.215

Notes: The OLS regression shows the interaction effect of the *positive* and *active treatment* group on the assessment of the 2°C goal, controlling for posterior returns, time, and number of clicks. *2°C assessment* reflects participants' z-scored evaluation of *alignment with the 2°C goal* relative to *exclusion of controversial companies* and *low emission intensity*. The variable *positive* is one if the subject learns about a positive correlation between *alignment with the 2°C goal* and (expected) returns and zero otherwise. *active* is an indicator variable that is one if a subject is part of the group that can actively choose the sustainability definition before the random selection of funds and zero otherwise. *time ETF selection* measures the duration spent on the fund selection page, and this variable is winsorized at the 5% and 95% levels. The variables *# clicks learning* and *# clicks decision* represent the number of clicks on different sustainability criteria in the learning (fund selection page) and are winsorized at the 90% level. *delta time* measures the time difference between the learning and decision screens and is winsorized at the 10% and 90% levels. All time variables are winsorized on the 10% and 90% levels () In all specifications, we adjust for demographics and pre-treatment return expectations. Standard errors are displayed in parentheses. Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

Table 33: Ownership of ETFs aligned with the 2°C sustainability definition increases the assessment of the 2°C goal sustainability definition

	(1) 2°C assessment	(2) 2°C assessment	(3) 2°C assessment	(4) 2°C assessment
owns 2°C goal asset	0.516***	0.235**	0.537***	0.312**
	(0.172)	(0.114)	(0.158)	(0.156)
Constant	-0.220	0.0608	-0.281***	-0.0560
	(0.151)	(0.0790)	(0.0859)	(0.0816)
N	166	302	158	182
controls	no	no	no	no
split	positive	positive	negative	negative
adj. R^2	0.0375	0.0105	0.0273	0.00406

Notes: The OLS regression indicates that ownership of ETFs *Aligned with the 2°C goal* sustainability definition increases the assessment of the 2°C goal sustainability definition. *2°C assessment* reflects participants' evaluation of *alignment with the 2°C goal* relative to *exclusion of controversial companies* and *low emission intensity*. Higher values indicate a stronger conviction that the *alignment with the 2°C goal* is more convincing. Each measure's persuasiveness is elicited on a Likert scale from one to five, before the mean of the other two definitions is subtracted from the *alignment with the 2°C goal* score. The variable is z-scored. The variable *own 2°C goal asset* is an indicator that is one if a participant is aware that she owns an asset aligned with the 2°C goal definition and zero otherwise. Columns (1) and (2) exclusively evaluate participants of the *positive treatment*, while Columns (3) and (4) focus on the negative treatment. Furthermore, Columns (1) and (3) evaluate participants who are aware of the attributes of their sustainable assets, while Columns (2) and (4) compare those owning a 2°C goal asset with those not knowing the attributes of their sustainable assets. Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively.

Chapter 4

4 Identifying Sustainable Investors: Insights from Machine Learning

Fabian Braun[a][b]

August 2024

Abstract

This study identifies the defining characteristics of next-generation retail investors and examines the predictive power of behavioral traits, demographics, and portfolio features in identifying sustainable investor types. Using an incentivized online stock market game with 640 young, well-educated German participants, I employ K-Means Clustering to develop data-driven personas for sustainable and non-sustainable investors. Subsequent analysis using Random Forest, XGBoost, K-Nearest Neighbor, and Logistic Regression reveals that while personal characteristics such as financial literacy, education, and income only modestly predict sustainable investing, including portfolio characteristics significantly enhances model performance. These insights can help researchers, regulators, and financial institutions in identifying sustainable investors, with implications for regulatory frameworks, product development, and marketing activities.

Keywords: ESG, Sustainable Investors, K-Means Clustering, Random Forests, Machine Learning

[a]The author gratefully acknowledges research support from the Leibniz Institute for Financial Research SAFE and the IPP Mainz. I thank Jonas Schwarze, Lea Voigt, and Max Herbert for their outstanding research assistance and seminar participants at Leibniz Institute for Financial Research SAFE Frankfurt, Goethe University Frankfurt, and Johannes Gutenberg University Mainz for helpful comments and discussions. The author declares no conflict of interest.

[b]Johannes Gutenberg University Mainz; Jakob-Welder-Weg 9, 55128 Mainz, Germany. Phone: +49 6131 3927901. E-Mail: braun@uni-mainz.de

4.1 Introduction

Sustainable Finance aims to use the power of financial markets to foster the green transition through affecting capital allocation. Given that policymakers seek to finance the green transition using private funds (Commission, 2019a) and that investors have preferences for sustainable products (Barreda-Tarrazona *et al.*, 2011; Riedl and Smeets, 2017; Bauer *et al.*, 2021; Braun *et al.*, 2024b), this necessitates the development of financial products that cater to these needs. Europe is currently the largest market for sustainable open-end and exchange-traded funds, holding 2,513 billion USD in Assets Under Management (AUM), which is 22% of all such funds in Europe (Morningstar, 2024a,b).[46] The market for sustainable products is not only substantial but potentially very profitable, as sustainable investors are willing to pay higher fees than conventional investors (Heeb *et al.*, 2023; Baker *et al.*, 2022; Engler *et al.*, 2023), and advisors and investment firms capitalize on this (Baker *et al.*, 2022; Laudi *et al.*, 2023). Understanding the characteristics of sustainable investors is therefore interesting for researchers, regulators, and financial institutions. Recent literature has identified several factors potentially influencing sustainable investments such as risk and return expectation, gender, altruistic preferences, age, financial literacy, income, and education (Bauer and Smeets, 2015; Døskeland and Pedersen, 2016; Riedl and Smeets, 2017; Gutsche *et al.*, 2019; Bauer *et al.*, 2021; Anderson and Robinson, 2022; Giglio *et al.*, 2023; Montagnoli and Taylor, 2024; Filippini *et al.*, 2024). However, what defines a typical sustainable investor and whether behavioral characteristics, demographics, and portfolio features can predict sustainable investor types remain open questions.

In this paper, I explore these questions among well-educated young adults in Germany who are about to start their first permanent employment and have either recently started or may soon begin investing in the stock market.[47] Given their potential for future high earnings, these individuals represent a significant target for banks and financial firms (DESTATIS, 2023a). Adopting an exploratory approach, I utilize K-Means Clustering to generate data-driven investor personas that characterize both sustainable and non-sustainable investors. Such personas are widely used to gain insights into the interests, thoughts, and behaviors of potential clients during product development or marketing campaigns (McGinn and Kotamraju, 2008; Miaskiewicz and Kozar, 2011; Akre *et al.*, 2019; Nielsen, 2019). If tailored products exist for different target groups, offering the appropriate products to the correct clients is beneficial as this reduces transaction costs, which may subsequently increase overall welfare (Häubl and Trifts, 2000). To elucidate what predicts sustainable investor types, I split the dataset into training and testing sets, train machine learning models on the training data, and subsequently predict the likelihood of sustainable investment behavior in a previously separated test dataset. The results show that while

[46] The Morningstar sustainable fund universe includes open-end funds and ETFs that, according to prospectus or other regulatory filings, focus on sustainability; impact; or environmental, social, and governance factors. This definition emphasizes intentionality over holdings to exclude funds that coincidentally hold sustainable assets.

[47] According to DAI (2024), most individuals in Germany start investing in stocks in their 20s. A recent survey indicates that 15.3% of the age group 19-29 holds stocks, which increases to about 20.5% for those aged 40-49.

personal characteristics such as financial literacy, education, and income only modestly predict sustainable investor types, portfolio characteristics significantly enhance model performance.

To elicit revealed preferences for sustainable investments, 640 individuals from the Universities of Mainz and Frankfurt (Germany) participated in an incentivized online stock market game in September 2021. They are presented with an investment opportunity involving 20 different real stocks, whose names are masked to prevent the use of external information. They receive comprehensive data on each stock, including past performance, dividend payment, P/E ratio, volatility, stock price, equity value, number of employees, turnover, debt ratio, and ESG Risk Scores, presented in a trading desk interface. To ensure incentive compatibility and to simulate a realistic investment environment, participants receive payouts from one of the portfolios constructed during the study after the session is concluded. Participants are also informed that real shares are purchased on the stock market based on their selections, ensuring that they bear the moral consequences of their investment choices. Additionally, the study integrates incentive-compatible measures of individual altruism (Andreoni, 1989; Crumpler and Grossman, 2008; Tonin and Vlassopoulos, 2010) and risk preferences (Sutter *et al.*, 2013), as well as survey measures of financial literacy (Hastings *et al.*, 2013; Lusardi and Mitchell, 2008), expectations about the risks and returns of sustainable investments (Riedl and Smeets, 2017), alongside with demographic data.

This investigation presents two principal findings. Firstly, I demonstrate that K-Means Clustering is an effective tool for constructing data-driven investor personas. Specifically, the analysis identifies two distinct personas: a sustainable investor and a non-sustainable investor. However, while these personas provide valuable information, it is crucial to note that the generated clusters partially overlap and a significant number of individuals could also fit into the alternate cluster. The first persona is a young female student from Mainz with an average pro-social attitude. Earning less than her peers and currently not owning any financial products, her financial education is slightly above average. Having recently learned about ESG investing, she expects these investments to provide moderately higher risk-adjusted future returns than conventional products. She prefers stocks from slightly larger-than-average companies that offer above-average dividends, high price-earnings ratios, above-average past returns, and low volatility. Lastly, the first persona prefers sustainable investments. Conversely, the second persona, a young male from Frankfurt University, shows a below-average pro-social attitude. His financial literacy is significantly higher than average. As a student with an above average income who already owns some stocks and funds, he is familiar with ESG products but does not anticipate risk-adjusted outperformance compared to conventional investments. He favors stocks from moderately above-average-sized companies that have slightly above-average short-term performance and excellent long-term returns. His stocks typically have average volatility, dividends, and price-earnings ratios. Lastly, this persona focuses mainly on financial products with low ESG Scores.

The second main finding relates to the use of Random Forests models to predict sustainable investors. The results demonstrate that personal characteristics possess some limited ability to predict sustainable investor types. Enriching the dataset with portfolio characteristics significantly increases the predictive power of the model. These findings are of particular significance for financial institutions and researchers seeking to identify sustainable investors. To address the issue of a mechanical correlation between ESG Scores and other portfolio characteristics, I focus on the four most salient stock characteristics in the main analyses and conduct several robustness tests. The evidence indicates that the results are unlikely to be driven by a mechanical correlation. This paper leverages SHAP (SHapley Additive exPlanations) (Lundberg and Lee, 2017) to give insights into the "black box" of machine learning, revealing that among personal characteristics financial literacy, education, and age are the most significant predictors. When portfolio characteristics are also considered, share price, revenue, and past returns emerge as the top predictors. The results are not driven by model choice. Robustness analyses including XGBoost, K-Nearest Neighbor, and Logistic Regressions confirm that Random Forest is an adequate model choice for the prediction task at hand.

This study introduces several key methodological distinctions compared to prior research. I specifically target young individuals, a group known to be starting to invest in Germany (DAI, 2024), highlighting its practical relevance to current and future market trends. Moreover, unlike previous studies which predominantly identify correlations between behavioral characteristics, demographics, and sustainable investments (Bauer and Smeets, 2015; Døskeland and Pedersen, 2016; Riedl and Smeets, 2017; Gutsche *et al.*, 2019; Bauer *et al.*, 2021; Anderson and Robinson, 2022; Giglio *et al.*, 2023), this research employs machine learning algorithms to generate investor personas and predict sustainable investors. I use K-Means Clustering to generate semi-fictional but insightful portrayals of potential investors, illuminating their thought processes and likely actions (Akre *et al.*, 2019). Researchers and regulators can use these findings to understand the characteristics of sustainable investors and adopt regulations accordingly. In addition, it facilitates financial institutions' customization of financial products to diverse client profiles. Furthermore, the data-driven clustering approach allows the data itself to determine the number of distinct investor personas, offering a flexible understanding of investor diversity. This contrasts with methods that yield predefined, often binary categorizations of investors as either sustainable or non-sustainable.

Moreover, to the best of my knowledge, this is the first scientific study to use machine learning methods to predict (sustainable) investor types. I evaluate the predictive capabilities of several models and find that the Random Forest algorithm is an adequate model choice for this purpose. This approach offers numerous advantages: First, Random Forests can capture nonlinearities and interactions, potentially revealing predictive variables that linear methods like OLS or Logistic Regressions might miss (Fuster *et al.*, 2022). Second, Random Forests are adept at managing large sets of features, including those that are irrelevant or highly correlated, thus reducing the risk of overfitting compared to traditional methods (DeMiguel *et al.*, 2023). Third, the existing literature does not predict out-of-sample or on unseen data (Riedl and Smeets, 2017; Faradynawati and Söderberg, 2022; Gutsche *et al.*, 2023), which poses the risk that the model does not generalize well on different datasets but even within the sample their estimations demonstrate a rather low goodness of fit. In contrast, I divide the dataset into training and test data and analyze the model performance on the separated test data.

Overall this paper contributes to the growing literature on sustainable investment behavior (Riedl and Smeets, 2017; Hartzmark and Sussman, 2019; Heeb *et al.*, 2023; Giglio *et al.*, 2023; Braun *et al.*, 2024b) by introducing investor personas and predicting sustainable investor types. Key studies suggest risk and return expectations of sustainable investments as one determinant for sustainable investment decisions (Døskeland and Pedersen, 2016; Hartzmark and Sussman, 2019; Dong *et al.*, 2022; Gutsche *et al.*, 2023; Giglio *et al.*, 2023; Braun *et al.*, 2024b). For instance, Døskeland and Pedersen (2016) demonstrate through a natural field experiment in an online banking context that financial considerations significantly motivate investors to opt for sustainable stocks. This finding aligns with Giglio *et al.* (2023), who survey thousands of Vanguard clients and conclude that although investors typically anticipate sustainable investments to underperform the market, there exists significant heterogeneity in these expectations. Crucially, only investors anticipating higher returns hold meaningful amounts of ESG assets. Contrarily, Riedl and Smeets (2017) link administrative data and survey responses from Dutch investors and conclude that the primary motivation for sustainable investors is not financial gain but social preferences.

Degryse *et al.* (2023) subsume such confronting views by showing that sustainable investors are divided between those investing for social reasons and those who believe in the financial outperformance of sustainable stocks. Additional factors potentially influencing sustainable investment decisions include gender (Junkus and Berry, 2010; Anderson and Robinson, 2022; Gutsche *et al.*, 2023), age (Junkus and Berry, 2010; Riedl and Smeets, 2017; Haber *et al.*, 2022), education (Junkus and Berry, 2010; Anderson and Robinson, 2022; Montagnoli and Taylor, 2024), and income or wealth (Haber *et al.*, 2022; Gutsche *et al.*, 2023), as well as financial literacy (Anderson and Robinson, 2022; Gutsche *et al.*, 2023; Montagnoli and Taylor, 2024; Auzepy *et al.*, 2024; Filippini *et al.*, 2024). However, the influence of these characteristics is not fully established. Contradictory evidence exists, with some studies showing little or no effect of these factors on sustainable investing (Riedl and Smeets, 2017; Døskeland and Pedersen, 2021; Haber *et al.*, 2022; Anderson and Robinson, 2022; Gutsche *et al.*, 2023).

This paper further contributes to the literature on predicting investment behavior with machine learning models (Shimokawa *et al.*, 2009; Chen *et al.*, 2019; Silva *et al.*, 2019; Rohatgi *et al.*, 2022; Mavruk, 2022). For example, demographics can be used to predict fraudulent investment behavior (Lokanan and Sharma, 2022), saving patterns and investment preferences (Rohatgi *et al.*, 2022), and risky stock trading behavior (Kim *et al.*, 2020). This paper contributes that demographics have some value in predicting sustainable investment behavior and enriching the dataset with portfolio characteristics greatly enhances the predictive capabilities of the model.

Lastly, this paper relates to the literature on the data-driven creation of investor personas (Cooper, 1999; McGinn and Kotamraju, 2008; Cooper *et al.*, 2014; An *et al.*, 2018; Nielsen, 2019; Salminen *et al.*, 2020; Jansen *et al.*, 2022). McGinn and Kotamraju (2008) highlight that data-driven approaches allow for the natural formation of groups within the data, from which personas can be derived. These personas provide valuable insights into the interests, thoughts, and behaviors of potential investors, thus facilitating a better understanding of who sustainable investors are. This enables more effective product development, targeted marketing campaigns, and better regulations. (Nielsen, 2019; Akre *et al.*, 2019). K-Means Clustering is a technique previously employed in the insurance and transportation literature to create data-driven personas (Arian *et al.*, 2021; Kumar and Oommen Philip, 2022). In this study, I contribute to this literature and show that Principal Component Analysis (PCA) and K-Means Clustering can be used to generate investor personas, thereby bridging the quantitative persona literature with sustainable investing practices.

The remainder of this paper is organized as follows: Section 4.2 provides an overview of the experimental design, data collection methods, and key variables. Section 4.3 details the main results from the clustering process and the generation of investor personas. In Section 4.4, I examine the predictive power of personal and portfolio characteristics. Section 4.5 discusses the findings and concludes the paper.

4.2 Experimental Design, Data Collection, and Key Variables

At the core of this paper is an incentivized stock market game wherein participants are tasked with assembling a portfolio from an array of real-world stocks. Each investor is provided with a trading account to purchase up to five different stocks from a selection of 20 (see Braun *et al.* (2024b) for more details[48]). The remainder of this section offers a brief description of the experimental design and an overview of the data collection process.

[48] The experiment was conducted jointly with Prof. Dr. Andrej Gill and Prof. Dr. Florian Hett from Johannes Gutenberg University Mainz. Research assistants selected 100 listed companies from Yahoo Finance according to four criteria: 1) real companies, no funds; 2) fiscal year ending in December; 3) companies categorized within five ESG Risk Score ranges: 0-10, 10-20, 20-30, 30-40, 40+; 4) classification into small/mid cap, large cap, and mega-cap categories. From the list of firms, 20 companies were randomly selected for the final sample, ensuring proportional representation from each ESG Risk Score bin and market cap bin. For the largest market cap bin, a sufficient number of sustainable firms was not available; hence, firms from the next smaller market cap bin were included.

Measurement of sustainability preferences: To measure investors' revealed sustainability preferences, we employ an incentivized stock market game. This methodology enables comprehensive control over the information accessible to participants and emphasizes real decision-making. Unlike survey responses or hypothetical scenarios, this approach mitigates the risk of hypothetical bias (List and Shogren, 1998; List, 2001; Harrison, 2006; Harrison *et al.*, 2008) and reduces the potential influence of experimenter demand effects (Haaland *et al.*, 2023) by introducing real consequences to participants decisions.

For the experiment, we use the prices and attributes of real shares; we mask the names of the corresponding shares and introduce an experimental currency (points). For instance, if the market price of a share of Adidas is 50 euros, its purchase price in the experiment is 50 points. Should the value of the Adidas share increase by 10% within one year on the actual stock exchange, its value in the experiment will similarly increase by 10%. Figure 1 shows a screenshot of the trading desk. When participants hover their cursor over one of the key figures, a brief explanation of the indicator is displayed. Clicking on the name of a stock opens a fact sheet presenting ten key figures for that share (see Figure 17 (appendix) for an example).

Account Balance:
50.000 Punkte

Name	Share Price (2019)	Past Performance (2018-2019)	P/E Ratio (2019)	Revenue (2020)	ESG Risk Scores (2020)
Firma 56	100,76	-17,56%	11,55	45.192.000.000	39,3
Firma 80	37,28	6,77%	29,39	31.303.440.000	25,1
Firma 7	12,92	4,78%	28,08	80.185.000.000	16,8
Firma 27	19,71	-15,00%	8,89	76.644.960.000	28,0
Firma 76	36,10	0,09%	8,06	3.160.200.000	7,8
Firma 94	19,02	6,07%	14,07	8.585.892.000	16,0

Figure 12: Trading desk

Notes: This figure presents the (translated) trading desk interface used by participants during the stock market game. Each participant is endowed with 50,000 points and can allocate this endowment among five different stocks selected from an assortment of 20. Initially, participants view key figures for each stock such as Share Price, Past Performance, Price Earnings Ratio, Revenue, and ESG Risk Score. By clicking on the stock's name, participants receive additional information, as illustrated in Figure 17, and have the option to purchase the stock.

Participants are displayed publicly available information on Past Performance, Dividends, Price Earning ratios, Volatility, and the Stock Price as of January 2019. To address concerns related to the efficient market hypothesis, we also provide data from 2020 on the value of Equity, the number of Employees, Revenues, Debt Ratio, and ESG Risk Score. Each participant is endowed with a trading account containing 50,000 points to purchase up to five different stocks at January 2019 prices from an assortment of 20 stocks, using the buy button located above and below the fact sheet.[49] Participants decisions result in two distinct measures of sustainability: The first measure (*ESG Score*) is determined by calculating the average ESG Score of the selected five stocks.[50] The second sustainability measure (*# Sustainable d.*) is an indicator variable that equals one if the portfolio includes more than two sustainable stocks.[51]

Measurement of personal characteristics: The literature identifies altruistic preferences, risk and return expectation, gender, age, financial literacy, income, and education as factors influencing SRI practices (Bauer and Smeets, 2015; Døskeland and Pedersen, 2016; Riedl and Smeets, 2017; Gutsche *et al.*, 2019; Bauer *et al.*, 2021; Anderson and Robinson, 2022; Giglio *et al.*, 2023; Montagnoli and Taylor, 2024; Filippini *et al.*, 2024). To elicit incentive-compatible measures of altruism, we implement two types of dictator games associated with a charity. In these games, participants allocate eight euros between themselves and a previously selected charity (Crumpler and Grossman, 2008; Tonin and Vlassopoulos, 2010). The first game is a standard dictator game, whereas in the second, known as a money-burning dictator game, participants are informed that the charity will receive exactly eight euros regardless of their donation decision. *Warm Glow Altruism* (satisfaction from doing good - giving by others is not a perfect substitute (Andreoni, 1989, 1990)) is measured by the donation amount in the money-burning dictator game, and *Impact Altruism* (utility from the benefit of others - impact altruists treat giving of others as perfect substitute (Andreoni, 1989, 1990)) is calculated as the difference between the donations in the standard dictator game and the money-burning dictator game. Moreover, we elicit incentivized measures of risk aversion (Sutter *et al.*, 2013), survey measures of risk and return expectations (Riedl and Smeets, 2017), and financial literacy (Hastings *et al.*, 2013; Lusardi and Mitchell, 2008), as well as a set of demographic variables. See Table 36 (appendix) for detailed definitions of each variable.

Data collection, payment of participants and incentive compatibility: Data collection was conducted through an oTree programmed online experiment (Chen *et al.*, 2016) with 640 participants affiliated with Goethe University Frankfurt and Johannes Gutenberg University Mainz between September 6th and 30th, 2021.[52] The sample predominantly consists of young, well-educated individuals who will start their first permanent jobs within the next few years. Given that the objective is to create personas for potential investors, this sample composition is rather a feature than a bug. In Germany, where stock market participation is relatively low

[49] Participants always select exactly five stocks but are allowed to buy the same stock five times.

[50] It is important to note that in the experiment, participants see the raw ESG Risk Scores from Sustainalytics, where a higher risk score indicates lower sustainability. For salience reasons, I linearly transform the ESG Risk Scores into ESG Scores using the formula *ESG Score = (2*mean(ESG Risk Score) - ESG Risk Score)* for each portfolio decision. Thus, in this paper higher ESG Scores imply greater sustainability.

[51] Sustainable stocks are defined by an ESG Risk Score of less than 20 (color-coded in the fact sheet with grey and yellow). Given that 40% of the stocks in the experiment are classified as sustainable, random selection would, on average, result in two sustainable stocks.

[52] From the original sample, I exclude three individuals aged 40, 55, and 62 because they do not fit the next-generation client profile and two participants who were suspected of having participated twice. Furthermore, I impute 25 missing values for *Risk Averse* and 19 missing values for *Perceived Financial Literacy* and *Precision Financial Literacy* using scikit learns k-nearest neighbor imputation using all other characteristics. All results remain qualitatively the same if these individuals are included (excluded).

with around 18% of the population over 14 years, individuals typically begin investing in stocks during their 20s (DAI, 2024). Those with bachelor's or master's degrees often command higher incomes, making them attractive future customers for financial institutions and an interesting group for regulators wanting to foster sustainable investments (Bogan, 2008; DESTATIS, 2023a; DAI, 2024).

As the described experiment is part of a larger study, this paper focuses solely on the initial portfolio allocation.[53] The stock market game is incentivized: for every participant, we pay out the value of one of the portfolios as of January 2020 at an exchange rate of 1 point = 0.0002 euros. Additionally, participants are informed that we purchase the randomly selected portfolio on the real stock market and invest 20% of its actual value using private funds. This ensures that investment decisions made during the game directly influence the financing of real companies, thereby imposing moral responsibility on the participants for these investments. Each portfolio is held for 12 months. Moreover, one of the decisions from the risk preference task, the standard dictator game, or the money-burning dictator game is paid out to participants (and the corresponding charities) following the experiment's completion.

Tables 37 and 38 in the appendix provide summary statistics for all variables.[54] The median time spent in the experiment was 50 minutes, and the median earnings were 18 euros per participant. This corresponds to an hourly wage of 21.60 euros, substantially higher than the typical earnings of 10.91-12.68 euros per hour for a student assistant in 2021. We offer these high incentives to ensure that the trades in the stock market game accurately reflect the decisions participants would make in the actual stock market.

4.3 Using K-Means Clustering to Create Investor Personas

The initial step of this research aims to identify distinct investor types, a crucial process in understanding who are the sustainable investors. To depict typical investor types, I utilize personas — semi-fictional representations of potential investors that provide insights into the investors´ thought processes and probable actions (Miaskiewicz and Kozar, 2011; Nielsen, 2019; Akre *et al.*, 2019). This paper proposes the creation of data-driven personas to characterize various investor types. The personas characteristics will be derived using a comprehensive dataset including demographics, behavioral characteristics, and portfolio features. The method employed is K-Means Clustering, an unsupervised classification technique that identifies data clusters by maximizing intra-cluster similarity and inter-cluster dissimilarity (Sinaga and Yang, 2020). The literature recognizes K-Means as one of the oldest and most widely used clustering methods, originating

[53] Participants create eight portfolios in total, each selected with a 12.5% probability for payout but in this paper, I focus on the first portfolio allocation. Subsequent stages introduce three variations of the stock market game. While these stages are insightful for assessing investor reactions to incentives and information, they are not suitable for addressing the questions posed in this paper. Details on later stages are provided in Braun *et al.* (2024b).

[54] See Braun *et al.* (2024b) for a detailed overview on the summary statistics.

from the seminal works of Sebestyen (1962) and MacQueen (1967). K-Means Clustering has applications across diverse fields including finance (Ellul *et al.*, 2020; Patton and Weller, 2022; Dou *et al.*, 2021; Begenau and Siriwardane, 2024), marketing (Jacobs *et al.*, 2016; Bekkerman *et al.*, 2023; Dzyabura *et al.*, 2023), and in the creation of personas (Arian *et al.*, 2021; Kumar and Oommen Philip, 2022).

4.3.1 K-Means Clustering

The basic principle of K-Means Clustering is to assign observations in a dataset to one of k clusters in order to minimize the squared distance from each observation to its nearest cluster center (MacQueen, 1967). This goal is articulated through an objective function given by

$$W(S,C) = \sum_{k=1}^{K} \sum_{i \in S_k} \|x_i - c_k\|^2 . \tag{2}$$

Here, S represents a partition of the dataset into K non-overlapping clusters S_k, where each cluster is defined by its centroid c_k for $k = 1, 2, \ldots, K$. x_i are vectors in a M-dimensional feature space (Xu *et al.*, 2016).

The K-Means algorithm starts by randomly selecting k initial points as centroids. It then alternates between two steps:

1. Assign all observations to the nearest centroid based on Euclidean distance
2. Recalculate the centroids of each cluster based on their current members

This cycle is reiterated until the centroids stabilize and no longer change. The final output comprises the classifications of each sample and the centroid of each cluster. Ashabi *et al.* (2020) conduct an extensive literature review and identify several challenges for K-Means Clustering: Defining the optimal number of k, finding the initial cluster centers, handling noise, outlier, and correlated variables. In the next step, I explain each challenge and how I cope with it in the dataset at hand.

Noise, Outliers, and Correlated Variables: Yu *et al.* (2018) highlight that K-Means Clustering is significantly affected by outliers and noisy data. To mitigate this, I employ the Interquartile Range (IQR) method for outlier detection (Tukey, 1977). The IQR is defined as the difference between the 25 and the 75 percentile of a variable. This method identifies values as outliers if they fall more than 1.5 times the IQR below the first quartile or above the third quartile. Standard practice is followed by replacing outliers with the nearest non-outlier values.[55]

[55] Upon reviewing all outliers, I do not adjust the binary variables *Possess Bonds* and *Master* since they rarely take the value of one and I would therefore lose these variables. As a robustness test, instead of using the IQR, I identify outliers by analyzing means, medians, deciles, and standard deviations. Those variables identified as having outliers I winsorize on the 5% and 95%. The results are qualitatively the same (unreported specification).

Features with smaller variations disproportionately influence clustering assignments based on distance metrics like inertia (cluster sum of squares) (Yu *et al.*, 2018). Consequently, I normalize the data with a mean of zero and a standard deviation of one. Furthermore, incorporating correlated variables that do not reflect the true cluster structure can obscure genuine clusters (Brusco and Cradit, 2001). Additionally, the objective function inertia is not a normalized metric and therefore in very high-dimensional spaces tends to become inflated (curse of dimensionality). PCA effectively addresses both issues by reducing dimensionality and deriving principal components from linear combinations of the original variables (Ding and He, 2004). The first principal component has the highest possible variance and therefore explains the majority of the inertia of the dataset, with each subsequent component being orthogonal to the previous and capturing the maximum possible inertia (Abdi and Williams, 2010; Chapman *et al.*, 2023). I apply PCA under the condition that the principal components retain 90% of the variance from the original data, effectively reducing the number of features for K-Means Clustering from 34 original features to 17 principal components.

Determining the Optimal Number of Clusters: K-Means requires defining the number of k before the execution of the algorithm. The machine learning literature offers various approaches for selecting the optimal number of clusters, k. One of the most popular methods, as noted by Begenau and Siriwardane (2024), is to choose the K-Means model that achieves the highest Silhouette score, $\overline{s}_k$ (Rousseeuw, 1987). In this approach, each observation i in a fitted K-Means model is assessed for cluster compatibility. $a(i)$ represents the average Euclidean distance to all other points within the same cluster, and $b(i)$ denotes the average distance to all points in the nearest cluster. The Silhouette score for observation i is calculated as:

$$s_k(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{3}$$

This value, $s_k(i)$, ranges between -1 and 1, with values close to 1 indicating a strong match of observation i to its cluster and negative values indicating that an observation is assigned to the wrong cluster. The overall Silhouette score, $\overline{s}_k$, is determined by averaging $s_k(i)$ across all observations. The higher the score the better is the clustering. Figure 13 plots the number of k on the x-axis against the average Silhouette score on the y-axis. The graph shows that the highest Silhouette score is found if $k = 2$ (0.11), whereas the Silhouette score is slightly lower for $k = 3$ (0.10) and then deteriorates for k between three and ten to around 0.07.[56] Using the elbow method for robustness purposes, I come to the same conclusion that two is the optimal number of clusters.[57]

[56] As the task at hand is to generate personas, more than six personas ($k > 6$) are unfeasible to operate for financial service providers (Nielsen, 2019). Nonetheless, in unreported robustness tests, I further increase k until $k = 50$ with the result that the Silhouette score stays at around 0.06 to 0.07.

[57] The elbow method is a visual method, where the number of clusters k is plotted against the inertia (Xu *et al.*, 2016). The optimal number of cluster k is the point at which the graph forms an elbow.

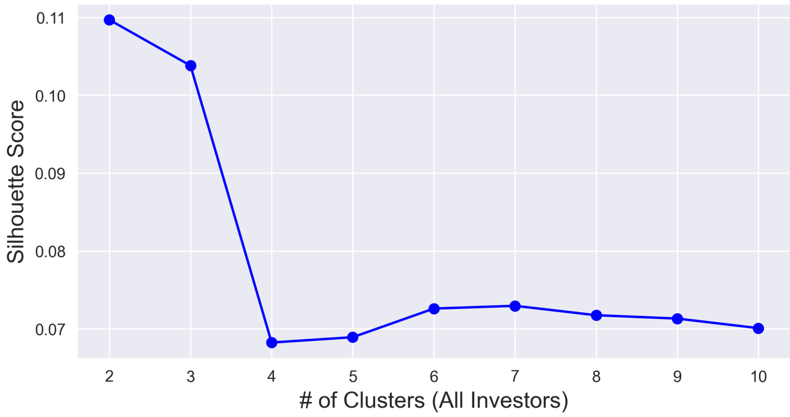


Figure 13: Optimal number of k

Notes: This Figure plots the number of k on the x-axis against the average Silhouette score on the y-axis. The graph shows that the highest Silhouette score is found if $k = 2$ (0.11), whereas the Silhouette score is slightly lower for $k = 3$ (0.10) and then deteriorates for k between three and ten to around 0.07. Thus, the optimal number of clusters is $k = 2$.

Initialization Problem: K-Means is a centroid-based clustering method that requires the pre-selection of initial cluster centers (Ashabi *et al.*, 2020). The necessity to randomly select k cluster centers can lead to local convergence due to the algorithm's greedy nature (Ikotun *et al.*, 2023). To mitigate this issue, I conduct 1,000 consecutive K-Means trials with random centroid seeds. The best result among these trials, determined by inertia, is considered the final outcome.

4.3.2 K-Means Clustering Results

As determined in the previous chapter, the Silhouette index indicates $k = 2$ as the optimal number of clusters for K-Means Clustering. After preprocessing the data (removal of outliers, standardization, and PCA), I fit a K-Means algorithm using the Python module scikit-learn (Pedregosa *et al.*, 2011) with 1,000 random initializations.[58] Table 34 presents the mean values for each cluster for all variables used as inputs for PCA. Additionally, it includes the p-values from the Mann-Whitney U test, which assesses differences between the groups. It is important to note that this analysis is purely descriptive and does not claim causality in any way. Cluster 0 comprises 248 individuals, while Cluster 1 contains 392 individuals.

Consequently, I generate two investor personas based on six supercategories derived from the data: (1) pro-social attitude, (2) (financial) education and experience, (3) risk and return expectations, (4) gender, university affiliation and other personal characteristics, (5) portfolio characteristics, and (6) sustainable investment preference.

[58] In a robustness test using K-Means++ instead of random initialization, the clusters remain largely the same and the Silhouette score decreases marginally (unreported test). According to Bahmani *et al.* (2012), in the K-Means++ algorithm, only the initial centroids are chosen randomly. Subsequent centroids are selected with a probability proportional to their squared distance from the nearest existing centroid. This method leverages the notion that effective clustering generally disperses the centers. Thus, when choosing new cluster centers, the algorithm preferentially selects points that are farther from the previously chosen centers, enhancing the likelihood of more distinct and well-separated clusters.

Table 34: Average values for the resulting two clusters of K-Means Clustering

	0	1	p-value (0 vs 1)
Age	25.41	23.15	0.00
Income	2.50	1.69	0.00
Possess Stocks	0.71	0.12	0.00
Possess Bonds	0.17	0.03	0.00
Possess Funds	0.72	0.16	0.00
Agrees with Organisation	5.81	5.96	0.07
ESG Knowledge	0.66	0.24	0.00
ESG Risk Lower	4.55	4.97	0.04
ESG Return Lower	3.72	3.85	0.68
Female	0.33	0.82	0.00
Has Siblings	0.42	0.53	0.01
Impact Altruism	1.69	2.00	0.07
Warm Glow Altruism	1.11	2.05	0.00
Risk Averse	0.47	0.47	0.29
# Financial Literacy	8.54	6.58	0.00
Perceived Financial Literacy	7.88	5.13	0.00
Precision Financial Literacy	0.43	0.30	0.00
Bachelor	0.42	0.21	0.00
Master	0.19	0.03	0.00
Other Education	0.40	0.76	0.00
Sample Mainz	0.28	0.65	0.00
Return 1 Month	-0.05	-0.05	0.93
Return 3 Months	-0.05	-0.02	0.00
Return 1 Year	0.03	0.05	0.09
Return 3 Years	0.79	0.68	0.02
Dividend	1.20	1.27	0.02
Price Earnings Ratio	15.15	16.58	0.00
Volatility	0.24	0.22	0.00
Share Price	36.74	36.93	0.87
Equity	19816255962.32	18660536245.55	0.12
# Employees	70645.04	67485.39	0.24
Revenue	23921449926.19	22736391134.69	0.20
Debt Equity Ratio	2.53	2.57	0.35
ESG Score	21.49	23.86	0.00
Number of Observations	248	392	

Notes: This table presents mean values for the resulting clusters of the K-Means Clustering with k=2 for all variables used as inputs for PCA. P-values are the results of a Mann-Whitney-U-Tests which assesses differences between the clusters.

Pro-Social Attitude: Investors in Cluster 1 exhibit an average pro-social attitude. They donate 25% of their endowment in a donation game that offers no tangible impact, and 50% when their donation has a direct impact.[59] Generally, they agree strongly with the organizations they donate to. Conversely, individuals in Cluster 0 show well below average pro-social attitudes, donating around 14% in the no-impact dictator game and 35% of their endowment when their

[59]Umer *et al.* (2022) conduct a meta-analysis and state that on average participants give 40-60% of their endowment in standard dictator games with charities. In comparison participants in Crumpler and Grossman (2008) give on average 20% of their endowment in money-burning dictator games with a charity as the recipient.

contributions have an impact. They also generally agree with the donation organizations. Taken together, investors in Cluster 1 are more altruistic than those in Cluster 0 (*Impact Altruism* ($p = 0.07$); *Warm Glow Altruism* ($p < 0.01$)), and more likely to agree with the organizations they support (*Agrees with Organisation* ($p = 0.07$)).

(Financial) Education and Experience: Investors in Cluster 1 are predominantly young students, either pursuing their bachelor's or master's degrees, with relatively low income compared to other students in Germany (Statista, 2021). A small fraction owns bonds, stocks, or funds, and their familiarity with the concept of ESG was limited prior to this study. Additionally, their financial literacy is slightly above average[60], and they underestimate their financial knowledge. In contrast, individuals in Cluster 0 are students in their mid-20s and well-educated, most holding already university degrees and being familiar with the concept of ESG. This group has a higher income than their peers (Statista, 2021) and predominantly owns bonds, funds, or stocks. They exhibit a well above-average financial literacy and overall are conscious of their financial knowledge. In summary, investors in Cluster 0 are more experienced with financial products, have attained higher levels of education, are slightly older, and possess higher incomes than those in Cluster 1 (*Possess Stocks, Possess Bonds, Possess Funds, # Financial Literacy, Perceived Financial Literacy, Precision Financial Literacy, Bachelor, Master, Other Education, Age, Income*; $p < 0.01$).

(ESG) Risk and Return Expectations: Investors in Cluster 1 expect slightly higher risk-adjusted returns for ESG assets compared to conventional assets. They perceive the risk of ESG assets as significantly lower compared to conventional assets, while they expect sustainable assets to slightly underperform conventional ones. On average, these individuals are almost risk-neutral. Conversely, investors in Cluster 0 perceive both the risk and returns of ESG assets to be slightly lower than those of conventional assets and therefore expect no positive risk-adjusted returns. Furthermore, they also display a near-risk-neutral attitude on average. Notably, the ESG risk expectations of investors in Cluster 1 are significantly lower than those in Cluster 0 ($p = 0.04$).

Gender, University Affiliation and Other Personal Characteristics: Investors in Cluster 1 are predominantly female, affiliated with Mainz University and half of them have siblings. In contrast, most investors in Cluster 0 are male, affiliated with Frankfurt University, and slightly less than half of them have siblings (*Sample Mainz*, *Female*, *Has Siblings*; $p < 0.01$).

[60]Compared to a representative Dutch sample (van Rooij *et al.*, 2011) where participants answered on average 54% and participants of a study with a German fintech company (Hett *et al.*, 2022) where participants answered on average 63% of largely similar questions correctly. Investors of Cluster 1 answered 66% and investors of Cluster 0 answered 85% of the questions correctly.

Portfolio Characteristics:[61] Investors in Cluster 1 hold stocks with comparably high past returns. Specifically, while their one-month performance is median, their three-month returns rank in the top 20%, and their one-year performance is in the top 30% of the available stock universe. In terms of long-term returns, their performance exceeds 80% of the available stocks. Additionally, their portfolio dividends are among the top 40%, their price-earnings ratios exceed 65% of the universe, and their volatility is below the median. Moreover, the size of the portfolio companies is among the largest 40% of the companies. Conversely, investors in Cluster 0 also prioritize past performance, maintaining median one-month returns, top 30% three-month returns, and top 35% one-year returns. Their three-year performance surpasses 85% of available stocks. Their portfolio dividends and volatility are around the median, while price-earnings ratios are among the top 40%. The size of the companies in their portfolios also ranks among the largest 40%. Overall, both groups exhibit high past returns, but Cluster 1 favors stocks with higher short-term returns (*Return 3 Months* ($p < 0.01$); *Return 1 Year* ($p = 0.09$)), whereas Cluster 0 prefers higher long-term returns (*Return 3 Years* ($p = 0.02$)). Additionally, Cluster 1's portfolios feature higher *Dividends* ($p = 0.02$), higher *Price Earnings Ratios* ($p < 0.01$), and lower *Volatility* ($p < 0.01$).

Sustainable Investment Preference: Investors in Cluster 1 manage portfolios that are considered sustainable. In contrast, individuals in Cluster 0 invest significantly less in sustainable assets, holding portfolios that are not deemed sustainable (*ESG Score*; $p < 0.01$).

Figure 14: Characteristics of the personas derived from Table 34

Notes: This figure displays the personas derived from the clustering analysis with $k = 2$ (see Table 34). The persona on the left (persona A) depicts a young female sustainable investor from Mainz. The persona on the right (persona B) shows a young male conventional investor from Frankfurt.

[61] Table 39 (appendix) shows the distribution of the universe of available stocks. I compare the actual values of the constructed portfolios to this distribution to get an approximation of the active portfolio choices made by participants.

The findings from Cluster 1 are personified in persona A (Figure 14, persona on the left). Persona A is a young female student from Mainz with an average pro-social attitude. Currently earning less than her peers, she does not own any financial products, yet her financial education is slightly above average compared to the general population. Recently, she has become acquainted with the concept of ESG investing and expects these investments to yield moderately higher risk-adjusted returns than conventional products. Persona A expects her future assets to have well above average short-term and long-term past returns. She has a preference for stocks of slightly larger-than-average companies that offer above-average dividends and high price-earnings ratios, with low volatility and high ESG Scores being a critical factor in her investment decisions.

The persona resulting from Cluster 0 is persona B (Figure 14, persona on the right). Persona B is a young male affiliated with Frankfurt University, exhibiting a below-average pro-social attitude. His financial literacy is very high compared to the general population. As a student with an above average income, he already owns some stocks and funds. Persona B is familiar with ESG products but does not expect them to yield risk-adjusted overperformance compared to conventional products. He has a preference for stocks with slightly above-average short-term past performance and ensures that the long-term past returns are among the top 20% in the market. Persona B favors stocks from moderately above-average-sized companies that offer average dividends and volatility, along with slightly above-average price-earnings ratios. This persona primarily focuses on conventional financial products.

Result 1: K-Means Clustering is successful and results in two data-driven investor personas: a sustainable and a non-sustainable investor. Nevertheless, the generated clusters are partially overlapping. Therefore, a large number of individuals could also be assigned to the other cluster.

According to Nielsen (2019), the number of personas required for a project depends on the diversity within the user group and the project's objectives. Given that the Silhouette score for $k = 3$ is only marginally lower, at 0.10 (refer to Figure 13), and considering the diversity within one cluster, the data also supports the creation of three personas. Table 40 in the appendix presents the mean values from K-Means clustering for $k = 3$. Three personas are illustrated in Figure 18 (see appendix), constructed using the same principles as for $k = 2$. Cluster 1 for $k = 3$ closely resembles Cluster 1 from $k = 2$, and is detailed under persona A's description. The second persona, persona D from Cluster 2, is a young male student with many similarities to persona B from Figure 14. He exhibits a slightly below-average social attitude and possesses a very high level of financial education. He perceives the risk-adjusted returns of ESG products as comparable to conventional products and plans to pursue a conventional investment strategy. The third persona, persona C from Cluster 0, also a young male student, shares characteristics with persona B but differs in his expectations. Persona C has a well below-average pro-social attitude and a very high financial education. Unlike persona B, he expects a small risk-adjusted overperformance from ESG products and prefers sustainable investments for the future.

Result 2: The data also supports the creation of three personas. This configuration retains persona A and introduces two new male investor personas, each sharing many characteristics with persona B. However, one predominantly invests in sustainable assets, while the other opts for conventional investments.

4.4 Predicting Sustainable Investors

After identifying personas that characterize different investor types, the next step involves predicting these investor types. The sample includes two data sets: (i) personal characteristics (demographics and behavioral characteristics), and (ii) portfolio characteristics that reflect customers' portfolio choices. An efficient allocation of financial products reduces transaction costs and potentially increases welfare (Häubl and Trifts, 2000). For example, banks may want to run targeted advertising campaigns, such as online banking pop-ups. Higher conversion rates ultimately reduce the banks' costs and thus potentially lower the fees for sustainable products.[62] Random Forest algorithms are highly effective in prediction tasks and have been utilized in various financial areas including the prediction of ESG Scores of firms (D'Amato *et al.*, 2022), credit defaults (Fuster *et al.*, 2022), future returns (DeMiguel *et al.*, 2023), market responses to news (Fedyk, 2024), inflation (Medeiros *et al.*, 2021), and trading intensity (Bogousslavsky *et al.*, 2024). Moreover, Random Forests effectively handle irrelevant or highly correlated predictors, enabling the inclusion of multiple characteristics with a reduced risk of overfitting compared to OLS or Logistic Regressions (DeMiguel *et al.*, 2023). Additionally, they can exploit nonlinearities and interactions, uncovering predictability that may be overlooked by linear methods such as OLS, Logistic Regressions, or Lasso (Fuster *et al.*, 2022).[63]

4.4.1 How Random Forests Work

Random Forests, introduced by Breiman (2001), are an ensemble machine learning algorithm that combines multiple decision trees to reduce overfitting and increase model performance. Hence, understanding Classification and Regression Trees (CART) (Breiman *et al.*, 1984) is crucial. CART are nonparametric models that approximate unknown nonlinear functions through local predictions by recursively partitioning the covariate space (Breiman, 1996). Initially, CART selects the feature from the training data that best groups data into sustainable and non-sustainable investors (Breiman *et al.*, 1984). This process repeats until additional splits no

[62]In a perfect market, lower costs lead to lower fees for customers and thus higher customer utility. But the market for sustainable investments is far from perfect. Whether the bank actually reduces overall fees is speculative and beyond the scope of this paper.

[63]Traditional statistical models, such as OLS or Logistic Regressions, are capable of incorporating nonlinear terms and interactions. However, in contrast to machine learning models, these need to be prespecified and are typically employed with caution to avoid overfitting (Fuster *et al.*, 2022).

longer enhance group differentiation. The decision tree can then be applied to the test data to forecast investor types. The goal is to generalize rather than perfectly model the training data, thus avoiding excessive tree splits (Breiman *et al.*, 1984). To reduce overfitting, it is common practice to use measures like gini $G(\tau)$, which indicates leaf purity.[64] As the number of splits increases, the prediction accuracy improves, but so does the number of leaf nodes. The pruning criterion for a CART model T is thus defined as:

$$C(T) = \sum_{\tau=1}^{|T|} G(\tau) + \lambda |T|, \tag{4}$$

where $|T|$ represents the number of leaf nodes in the model, measuring decision tree complexity, and λ is a regularization parameter determined by cross-validation (Khandani *et al.*, 2010). The goal is to minimize $C(T)$, balancing between the number of splits and the penalty term to optimize tree complexity. The algorithm seeks a compromise between increased splits for higher leaf node purity and the penalty term.

Individual CART typically exhibit high variance and are prone to overfitting. To mitigate these issues, the Random Forest model employs bootstrap aggregation (bagging) of randomly constructed tree predictors (Breiman, 2001). This process involves creating multiple subsets of the training data through random sampling with replacement, which are then used to train an ensemble of decision trees. Each tree in the ensemble relies on features from a random vector sampled independently but with the same distribution across all trees in the forest. This method captures diverse patterns and relationships in the data, resulting in more accurate and stable predictions, particularly for high-dimensional datasets or those with many irrelevant features. The randomness introduced in forests results in trees with less correlated prediction errors, allowing some errors to cancel out when averaging predictions. To ensure a model's validity is not compromised by specific characteristics of a dataset absent in the general population, it is standard practice to split the dataset into training and test sets (Khandani *et al.*, 2010; Géron, 2018). The model is parameterized using the training data, while the test data serves to validate the findings on a previously unused dataset. For implementing the Random Forest, I utilize the scikit-learn package (Pedregosa *et al.*, 2011), which averages the probabilistic predictions of the classifiers.

Hyperparameters significantly influence the performance of machine learning algorithms (Bergstra and Bengio, 2012; Wu *et al.*, 2019). Hyperparameter optimization involves training a model on various combinations of parameters, evaluating performance on a cross-validation set, and selecting the configuration that yields the best performance (Wu *et al.*, 2019). For optimizing the Random Forest model, I utilize the SMAC3 package (Lindauer *et al.*, 2022), a Bayesian optimizer known for its efficiency and lower resource requirements compared to manual

[64] In the Random Forest, I also consider entropy and log loss as alternative measures. The optimal measure is determined through hyperparameter optimization.

search or grid search (Awad *et al.*, 2020).[65] The objective is to identify a robust configuration by evaluating less promising candidates on a few instances and then collecting more empirical evidence on many instances if the candidate shows potential. Following established practices, I aim to maximize the area under the receiver operating curve (ROC AUC) across various parameter settings. The ROC AUC value represents the probability that the model assigns a higher score to a randomly selected sustainable investor compared to a non-sustainable one, and is an often used metric in machine learning (Calders and Jaroszewicz, 2007; Yang and Ying, 2022).

The same features used for creating personas are employed to predict investor types. Summary statistics are displayed in Tables 37 and 38 (appendix). Due to potentially new or limited client relationships, researchers and banks may lack extensive portfolio data. Given that the target group in this study comprises well-educated young individuals soon to start their careers and begin investing, this assumption is appropriate. For this reason, predictions are based solely on personal characteristics in the first step; later, portfolio characteristics are also incorporated. A sustainable investor is defined, as detailed in Section 4.2, as an investor who holds more than two sustainable stocks.

4.4.2 Predicting Sustainable Investors Using Personal Characteristics

I train a Random Forest model on the training data using SMAC hyperparameter optimization and fivefold cross validation. The machine learning predictions result in a score, indicating the probability that a certain investor can be considered sustainably. Setting the appropriate level for the threshold, when a participant is labeled as sustainable and therefore offered a sustainable investment product, involves a trade-off. A very low threshold results in many investors being classified as sustainable investors, and while it may correctly identify investors who are genuinely sustainable investors, this choice may also result in individuals incorrectly classified as sustainable investors and therefore hurt the sales numbers of non-sustainable products. On the other hand, a high threshold may classify a lot investors as non-sustainable, while their true preference are sustainable products, resulting in lower sales of sustainable products.

Confusion Matrix: Table 35 shows the confusion matrix of the results. It consists of two rows corresponding to the actual investor types — sustainable (investors whose portfolios contain more than two sustainable stocks) and non-sustainable. The columns represent the predicted labels. To find the optimal threshold when to label investors as 'Sustainable' (Sus) and 'Conventional' (Con), I employ Youden's J statistic (Youden, 1950), a widely utilized summary measure of the Receiver Operating Curve (ROC) (Schisterman *et al.*, 2008). This statistic assumes equal costs

[65] See Wu *et al.* (2019) for a comprehensive discussion on the advantages of Bayesian optimizers for hyperparameter optimization, including comparisons to manual search, grid search, and random search methods. Hyperparameter optimization for the Random Forest model includes the number of trees, maximum tree levels, minimum samples to split an internal node, minimum samples at a leaf node, number of features at every split, and the function to measure split quality (gini impurity, log loss, and entropy).

for False Positives and False Negatives and is defined as follows:

$$YJS = \frac{\text{TP}}{\text{TP} + \text{FN}} + \frac{\text{TN}}{\text{TN} + \text{FP}} - 1 \tag{5}$$

I calculate Youden's J statistic for each threshold, selecting the threshold that maximizes this statistic, which results in an optimal threshold of 0.61. The test data consists of 117 Sustainable and 75 Non-Sustainable investors. Table 35 illustrates that the Random Forest model correctly identifies 55 of the truly sustainable investors (True Positives) and 50 of the truly non-sustainable investors (True Negatives). Simultaneously, it misclassifies 25 investors as sustainable who are truly non-sustainable (False Positives), and 62 investors as non-sustainable who are truly sustainable (False Negatives). This results in an overall model Accuracy of 54.69%. The model achieves a Precision of 68.75%, indicating the percentage of correctly identified sustainable investors out of all labeled as such, and a Recall of 47%, reflecting the percentage of actual sustainable investors correctly identified. There is an inherent trade-off between Precision and Recall, where efforts to maximize one metric typically reduce the other (Géron, 2018). The F1 Score, which balances Precision and Recall by their harmonic mean, is 55.84% for the machine learning model that relies solely on personal characteristics.

		Model Prediction	
		Sus.	Con.
Actual Outcome	Sus.	TP	FN
	Con.	FP	TN

Performance Metrics

Accuracy $= \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}$

Precision $= \frac{\text{TP}}{\text{TP+FP}}$

Recall $= \frac{\text{TP}}{\text{TP+FN}}$

F1 Score $= \frac{2x\text{Precision}x\text{Recall}}{\text{Precision+Recall}}$

		Model Prediction	
		Sus.	Con.
Actual Outcome	Sus.	55	62
	Con.	25	50

Performance Metrics

Accuracy = 54.69%

Precision = 68.75%

Recall = 47.00%

F1 Score = 55.84%

Table 35: Model performance for classification using personal characteristics

Notes: Confusion matrix framework and model performance metrics based on personal characteristics for the optimal threshold as determined via Youden's J statistic (0.61). Rows correspond to actual investor type and columns correspond to forecasted types. TN: True Negative, FN: False Negative, FP: False Positive; TP True Positive

Receiver Operating Curve and Area Under the Curve: The confusion matrix uses a fixed threshold to classify sustainable investors. This is often not the best approach, as the true cost of False Positives (Negatives) may be unknown. To provide an overview of results across various classification thresholds, the blue line in Figure 15 displays the ROC, with the False Positive Rate on the x-axis and the True Positive Rate on the y-axis for the model relying exclusively on personal characteristics. Ideally, one aims for the highest possible True Positive rate with the lowest False Positive rate. Thus, banks face a trade-off: aggressively marketing sustainable products requires a lower threshold, while marketing non-sustainable products suggests a higher threshold. The optimal choice depends on the opportunity costs associated with losing sustainable clients. The ROC illustrates that this trade-off is not linear. If the cost of False Positives is equal to the gain from True Positives, the optimal threshold corresponds to the point where the ROC curve tangentially meets the 45° line (Khandani *et al.*, 2010). If the trade-off differs from one-to-one, the optimal threshold is adjusted accordingly.

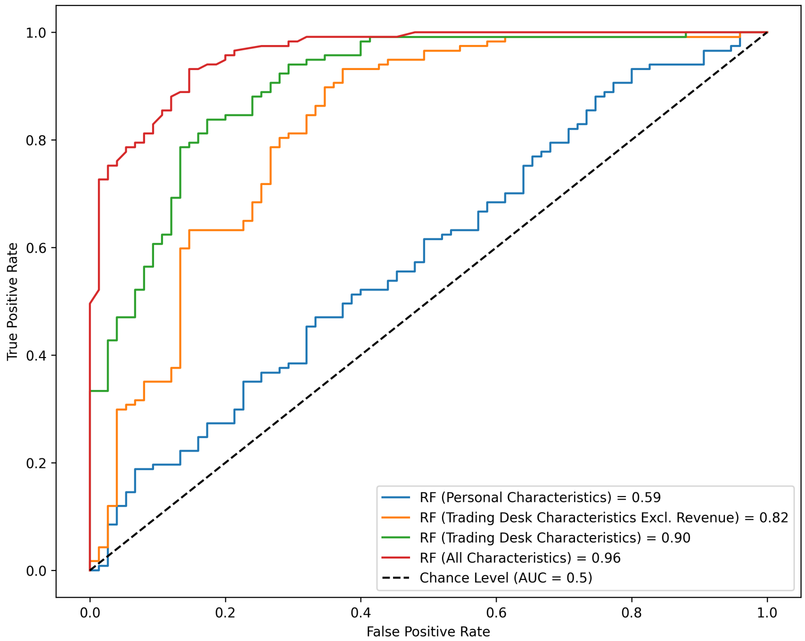


Figure 15: ROC for various Random Forest models

Notes: Figure 15 displays the receiver operating curve (ROC), with the False Positive Rate on the x-axis and the True Positive Rate on the y-axis for Random Forest Models based on four different sets of features: (i) personal characteristics alone, (ii) personal characteristics combined with portfolio characteristics visible on the trading desk excluding *Revenue*, (iii) personal characteristics combined with all portfolio characteristics visible on the trading desk, and (iv) all available personal and portfolio characteristics. The predictive power of the personal characteristics is rather low but better than chance level, while the model performance increases when portfolio characteristics are included.

The classifier's performance can be summarized by the ROC AUC, which is calculated by integrating the area under the ROC (Muschelli, 2020). A perfect classifier scores an ROC AUC of 1, while a random classifier scores 0.5 (Géron, 2018; Flach *et al.*, 2011). This measure is preferred for its straightforward interpretation, scale invariance, and independence from arbitrary classification thresholds (Bradley, 1997; Flach *et al.*, 2011).[66] Additionally, ROC AUC is effective with skewed class distributions, relevant as approximately 61% of investors in the sample are classified as sustainable (Calders and Jaroszewicz, 2007). The overall model performance of the model relying exclusively on personal characteristics is 0.59, thus slightly better than chance.[67]

[66]While ROC AUC is critiqued for using different misclassification cost distributions across classifiers (Hand, 2009; Muschelli, 2020), it provides a comprehensive overview of a classifier's performance without initial assumptions about the costs of false outcomes.

[67]The definition of good model performance varies significantly across applications and industries. As a general guideline, Hosmer *et al.* (2013) suggests that an ROC AUC value between 0.5 and 0.7 indicates poor discrimination between two classes, while a ROC AUC range from 0.7 to 0.8 denotes acceptable discrimination and an

Explaining the Machine Learning "Black Box" using SHAP: A notable drawback of Random Forests, compared to simpler machine learning models such as CART, is their "black box" nature due to the ensemble of multiple trees. In order to gain an understanding of which features are important for the model and increase trust of all stakeholders into the model, I use SHAP (Lundberg and Lee, 2017), a state-of-the-art explainable AI tool (Gramegna and Giudici, 2021). SHAP clarifies predictions by calculating each feature's contribution using coalitional game theory principles, where the feature values of a data instance act like players in a coalition (Molnar, 2022). Shapley values (Shapley, 1952) thus determine the fair contribution of each feature to the prediction outcome, employing linear models that weight instances according to the characteristics of these coalitions (Bauer *et al.*, 2023). In a nutshell, SHAP produces Shapley values that represent model predictions as linear combinations of a set of features, indicating whether each covariate is included in the model or not (Gramegna and Giudici, 2021). Positive values indicate an increase, while negative values indicate a decrease in the likelihood of being a sustainable investor. Please note that all effects described represent correlations within the model and do not imply causation.

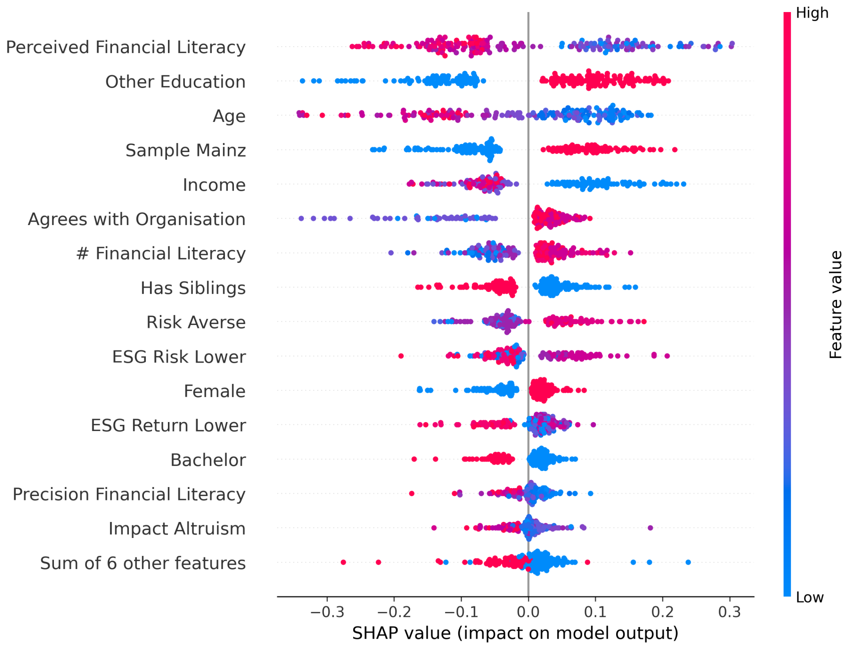


Figure 16: SHAP Beeswarm Plot (Random Forest model using personal characteristics)
Notes: This Figure represents the SHAP Beeswarm Plot for the Random Forest model using personal characteristics. Each point on the x-axis represents the SHAP value for an investor for a specific feature, with the color indicating the feature's value from low to high. All features are ordered by importance on the y-axis. The plot displays a negative correlation between being a sustainable investor and *Age, Perceived Financial Literacy, Income*, and *Bachelor*. Conversely, it shows a positive correlation for being a sustainable investor with *other education* and mixed evidence for *ESG Risk Lower*.

Figure 20 (appendix) illustrates the SHAP feature importance by averaging the absolute Shapley values per feature across the dataset (Molnar, 2022). Notably, *Perceived Financial Literacy, Other Education, Age, Sample Mainz*, and *Income* emerge as the most significant

ROCAUC > 0.8 indicates excellent discrimination capabilities of an algorithm.

features.[68] For a more granular analysis, I use the Beeswarm Plot in Figure 16. Each point on the x-axis represents the SHAP value for an investor for a specific feature, with the color indicating the feature's value from low to high. All features are ordered by importance on the y-axis. The plot displays a negative correlation between being a sustainable investor and *Perceived Financial Literacy, Age*, and *Income*. Conversely, it shows a positive correlation for being a sustainable investor with *Other Education* and *Sample Mainz*. Although this result should be interpreted with caution, it is noteworthy given the extensive literature identifying risk and return expectations (Døskeland and Pedersen, 2016; Riedl and Smeets, 2017; Hartzmark and Sussman, 2019; Dong *et al.*, 2022; Braun *et al.*, 2024b; Gutsche *et al.*, 2023; Giglio *et al.*, 2023) and altruistic preferences (Riedl and Smeets, 2017; Bauer *et al.*, 2021; Gutsche *et al.*, 2023) as key drivers of sustainable investments. In this dataset, these factors demonstrate quite low predictive power.

To summarize, personal characteristics exhibit some predictive power for identifying sustainable investor types, but overall, their effectiveness is limited. This finding highlights a significant gap in existing research, as there appears to be no successful application of machine learning methods for predicting sustainable investor types to date. Previous studies such as those by Riedl and Smeets (2017), Faradynawati and Söderberg (2022), and Gutsche *et al.* (2023) primarily use linear models to establish correlations between personal characteristics like demographics, altruism, and return expectations. However, these models do not predict out of sample and even within sample, their estimations demonstrate low goodness of fit.

Result 3: Personal characteristics possess some predictive power for identifying sustainable investors. However, overall, the predictive power is relatively low.

4.4.3 Predicting Sustainable Investors Using Personal Characteristics and Portfolio Choices

Portfolio choices might inform about the investment style of individuals. Hypothetically, non-sustainable investors may be characterized as growth investors, focusing on young companies experiencing above-average increases in revenues, profits, or cash flows. Consequently, they are likely to purchase stocks with higher price-to-earnings ratios and lower dividend yields. In contrast, sustainable investors might lean towards value investing, targeting older companies with higher book-to-market ratios and lower price-to-earnings ratios, which typically offer higher dividends (Cronqvist *et al.*, 2015; Betermier *et al.*, 2017). To explore investors' preferences for certain investment styles, I include the portfolio characteristics of the selected stocks in the machine learning model. However, a mechanical correlation might exist between stock characteristics

[68]The SHAP package used assumes feature independence. However, if features are correlated, this can result in unrealistic explanations (Aas *et al.*, 2021). Given that the objective of this paper is to predict investor types rather than provide causal explanations, I will not delve further into this issue.

and ESG Scores. For instance, although investors prioritize ESG Scores leading to selections of portfolios with high scores, the available stocks' ESG Scores might inherently correlate with past performance, influencing portfolios to also exhibit a high past performance mechanically. Figure 19 in the appendix shows the Spearman Rank Correlation among the 20 stocks from the original stock universe. It notably highlights a strong negative correlation between *Volatility* (*Revenue*) and *ESG Scores*, alongside a strong positive correlation between *Return 3 Months* and *ESG Scores*. To mitigate concerns of mechanical correlations, the analysis concentrates on the four stock characteristics most salient to investors because they are depicted on the trading desk (see Figure 12 for an example), while all other metrics are only visible if participants click on the stocks name. Among these, the Spearman correlations between *ESG Scores* and *Return 1 Year* (0.10), *Price Earnings Ratio* (0.15), and *Share Price* (0.01) are relatively low, with the exception of a strong negative correlation between *ESG Scores* and *Revenue* (-0.51).

Including portfolio characteristics significantly enhances the predictive power of the Random Forest model. As shown by the green line in Figure 15, the model's ROC improves across all thresholds. Consequently, the ROC AUC increases from 0.59, when only personal characteristics are considered, to 0.90 with the inclusion of portfolio characteristics. Additionally, the Accuracy, Precision, Recall, and F1 Score improve to 83%, 88%, 84%, and 86%, respectively (see Table 41, Panel A, in the appendix). The SHAP beeswarm plot in Figure 21 (appendix) reveals that the four newly added features are the most influential in explaining the predictions, with *Share Price* being the most impactful. The results demonstrate a negative correlation between being a sustainable investor and the features *Share Price*, *Revenue*, and *Return 1 Year* and a mixed relation between being a sustainable investor and *Price Earnings Ratio*.

To address concerns regarding a mechanical correlation between portfolio characteristics and *ESG Scores* in the original datset, I provide three pieces of evidence: First, the most important predictor in the Random Forest Model is *Share Price*, which is unrelated to *ESG Scores* in the original dataset (Spearman correlation of 0.01). If a purely mechanical correlation were driving the result, the most important predictor should be *Revenue* as it has the highest correlation in the original dataset (Spearman correlation of -0.51). Second, if anything, the Spearman correlations between *Price Earnings Ratio* (0.15), *Return 1 Year* (0.10), *Share Price* (0.01), and *ESG Scores* are positive in the original dataset. However, the SHAP values show a negative (mixed) relation between these variables and sustainable investment decisions. If the result were driven by a mechanical correlation, the relation should be positive. Third, *Revenue* has the highest correlation with *ESG Scores* in the original dataset (-0.51). Hence, if this feature mainly drives the demonstrated results, excluding it should decrease the ROC AUC to around the level of the model relying exclusively on personal characteristics. However, the results in Figure 15 (orange line) show only a minor reduction of the ROC AUC to 0.82. This is still substantially higher than the 0.59 achieved by the model based solely on personal characteristics.

Taken together, I find no evidence that the increased model performance due to the inclusion of portfolio characteristics is mainly driven by a mechanical correlation.

Further robustness specifications include all portfolio characteristics to predict sustainable investor types, which predictably boosts all performance metrics even further (Figure 15 (red line); Table 41 Panel A, appendix). The enhanced model achieves a ROC AUC of 0.96 and an overall Accuracy of 90%. Overall, portfolio characteristics demonstrate substantial predictive capability in this study.

Result 4: Adding portfolio characteristics substantially increases the predictive power of Random Forest models in identifying sustainable investor types.

As part of a robustness test, I employ XGBoost, K-Nearest Neighbor, and Logistic Regression models to compare their performance against the Random Forest Model.[69] The results, detailed in Table 41 Panels A, B, C, and D in the appendix, incorporate three different sets of features: (i) personal characteristics alone, (ii) personal characteristics combined with portfolio characteristics visible on the trading desk, and (iii) all available personal and portfolio characteristics. The Random Forest consistently ranks among the top-performing models, particularly excelling by achieving the highest ROC AUC for the dataset based solely on personal characteristics. Thus, it is an adequate model choice for the prediction tasks given the available dataset.

Result 5: The results are robust to model choice. Random Forest emerges as an adequate model for predicting investor types in the dataset at hand.

4.5 Discussion and Conclusion

The objective of this paper is to utilize machine learning algorithms to (i) develop data-driven personas for distinct investor types and (ii) predict sustainable investor types with varying information sources. To do so, the paper employs the most optimal machine learning models for the available data. The data utilized in this paper is derived from an incentivized online stock market game. I demonstrate the existence of at least two distinct investor personas: a sustainable and a non-sustainable. These personas are distinguished by variations in pro-social attitude, financial education and experience, ESG risk and return expectations, and preferences for various stock characteristics. However, the differences between the different investor types are often subtle, and there are investors who can be assigned to both personas. Furthermore, this paper demonstrates that personal characteristics alone have limited predictive power for identifying

[69]For the XGBoost model, I employ scikit-learn's grid search hyperparameter optimization technique to refine several parameters: learning rate, maximum depth of a tree, minimum loss reduction required to make a split (gamma), the fraction of observations and features randomly sampled for each tree (subsample and colsample bytree), and the minimum sum of instance weights needed at a child node. For K-Nearest Neighbor, I utilize the RandomizedSearchCV function from scikit-learn, optimizing for the number of neighbors, weights, the power parameter for the Minkowski metric, and leaf size. For Logistic Regression, I use scikit-learn's grid search hyperparameter optimization technique to refine the regularization parameter, the penalty term, the solver, and the maximum number of iterations.

sustainable investor types. However, this can be significantly enhanced by including portfolio characteristics. These results are not an artifact of model choice or mechanical correlation.

The findings of this paper have the potential to be of interest to researchers, regulators, and financial institutions. The results provide insights into the interests, thoughts, and behaviors of young investors. Researchers and regulators may utilize these insights to develop models and regulations that integrate these investor personas into their frameworks. Moreover, financial institutions may build upon this research to develop sustainable investment products that facilitate financing the green transition of the economy. The targeting of investors with the appropriate products may result in a reduction in transaction costs, thereby benefiting both financial institutions and their clients.

This paper makes use of correlational predictions. However, further research could employ causal machine learning models (e.g., see Athey *et al.* (2019)) to uncover causal relationships in predicting sustainable investor types. Moreover, personal characteristics may differ among young adults within the same age group. These unobservables might be correlated with sustainable investment behavior. Furthermore, it is possible that investors with higher stakes and in their usual online banking environment may behave differently.[70] Future studies should utilize representative samples and administrative (account) data to develop more precise investor personas. Another promising area for future research could be to leverage bank transaction data for the development of investor personas and predicting sustainable investors. In their study, Famulok *et al.* (2023) use account transaction data from a large German bank to calculate the carbon footprint of individuals. They find that higher footprints are linked to greener portfolios. In contrast, Qian *et al.* (2023) elicit pro-environmental behavior from account transaction data of Alipay clients. They observe a positive correlation between this pro-environmental behavior and sustainable investments. Finally, this study employs a data-driven approach to creating investor personas. Qualitative interviews (e.g., see Salminen *et al.* (2020)) can be used to validate these personas, thereby enhancing the depth and accuracy of the investor profiles derived.

[70] To conduct an optimal study, it would be beneficial to have administrative data from a bank with a representative sample of 20-30 year old clients. It is important to ensure that participants are unaware that they are being studied in order to reduce the Hawthorne effect (Roethlisberger and Dickson, 2003) and experimenter demand effects (Haaland *et al.*, 2023). Furthermore, the universe of stocks should not be limited to 20 stocks but should include a greater number of stocks with characteristics unrelated to *ESG Scores*.

4.6 Appendix

Table 36: Variable definitions

Variable	Measure
Age	The investor´s self-reported age.
Income	Investor´s answer to the question *"How high is your personal net income per month"*: (1; 0-700 euro) (2; 701 - 1,000 euro) (3; 1,001 - 1,500 euro) (4; 1,501 - 2,500 euro) (5; more than 2,500 euro)
Possess Stocks	Dummy variable equal to one if the investor holds stocks in the real stock market.
Possess Bonds	Dummy variable equal to one if the investor holds bonds in the real stock market.
Possess Funds	Dummy variable equal to one if the investor holds funds in the real stock market.
Agrees with Organisation	The participants' response to the question *"I completely identify with the goals of the organization"* from 0 to 7.
ESG Knowledge	Dummy variable equal to one if the investor knows the concept of ESG before the Experiment.
ESG Risk Lower	The investor´s response to the statement *"I expect stocks with low ESG risk scores to be subject to lower price fluctuations than stocks with high ESG risk scores"* on a scale from 0 to 7.
ESG Return Lower	The participant´s response to the statement *"I expect stocks with low ESG risk scores to have lower returns than stocks with high ESG risk scores"* on a scale from 0 to 7.
Female	Dummy variable that is one if an participant reports to be female.
Has Siblings	Dummy variable that is one if a participant reports to have siblings.
Impact Altruism	Difference of donation in impact and money burning dictator game [-8;+8].
Warm Glow Altruism	Donation in money burning dictator game.
Risk Averse	Indicates participants´s risk attitude as measured in the risk elicitation task. A higher value indicates that the participant is more risk averse [0;0.94].
# Financial Literacy	The number of correctly answered questions of the financial literacy quiz [0;10].
Perceived Financial Literacy	Sum of subjective probability weights * expected correct answers for each potential number of correct answers.
Precision Financial Literacy	Sum of squared subjective probability weights assigned to each potential number of correct answers.
Bachelor	Dummy variable equal to one if the participant´s highest educational qualification is a bachelor´s degree.
Master	Dummy variable equal to one if the participant´s highest educational qualification is a master´s degree.

Table 36 (continued):

Variable	Measure
Other Education	Dummy variable equal to one if the participant´s highest educational qualification is not a bachelor´s or master´s degree or higher (e.g. high school diploma etc.)
Sample Mainz	Dummy variable equal to one if the participant participated in the study via the Johannes Gutenberg University Mainz and zero if a participant participated via Goethe University Frankfurt.
Return 1 Month	The average 1-month past return of a participant´s stocks in the portfolio.
Return 3 Months	The average 3-month past returns of a participant´s stocks in the portfolio.
Return 1 Year	The average 1-year past return of a participant´s stocks in the portfolio.
Return 3 Years	The average 3-years past returns of a participant´s stocks in the portfolio.
Dividend	The average dividends of a participant´s stocks in the portfolio.
Price Earning Ratio	The average price earnings ratio of a participant´s stocks in the portfolio.
Volatility	The average volatility of a participant´s stocks in the portfolio.
Share Price	The average share price of a participant´s stocks in the portfolio.
Equity	The average equity value of a participant´s stocks in the portfolio.
# Employees	The average number of employees of a participant´s stocks in the portfolio.
Revenue	The average revenue of a participant´s stocks in the portfolio.
Debt Equity Ratio	The average debt equity ratio of a participant´s stocks in the portfolio.
ESG Score	The average ESG Score of an participant´s portfolio choice in portfolio allocation 1.
# sustainable d.	Dummy variable equal to one if the participant invests in more than two sustainable stocks in a given portfolio allocation.

Notes: Table 36 defines all relevant variables of the study.

Table 37: Summary statistics of personal characteristics

	count	mean	std	25%	50%	75%
Age	640	24.09	3.54	21.75	23.00	26.00
Income	640	2.00	1.16	1.00	2.00	3.00
Possess Stocks	640	0.35	0.48	0.00	0.00	1.00
Possess Bonds	640	0.08	0.28	0.00	0.00	0.00
Possess Funds	640	0.37	0.48	0.00	0.00	1.00
Agrees with Organisation	640	5.90	0.94	5.00	6.00	7.00
ESG Knowledge	640	0.40	0.49	0.00	0.00	1.00
ESG Risk Lower	640	4.78	1.76	4.00	5.00	6.00
ESG Return Lower	640	3.80	2.01	2.00	4.00	5.00
Female	640	0.63	0.48	0.00	1.00	1.00
Has Siblings	640	0.49	0.50	0.00	0.00	1.00
Impact Altruism	640	1.87	2.68	0.00	1.00	4.00
Warm Glow Altruism	640	1.71	2.23	0.00	0.00	3.00
Risk Averse	640	0.47	0.14	0.44	0.44	0.56
# Financial Literacy	640	7.33	1.97	6.00	8.00	9.00
Perceived Financial Literacy	640	6.20	2.22	4.80	6.26	7.93
Precision Financial Literacy	640	0.37	0.20	0.24	0.30	0.44
Bachelor	640	0.29	0.45	0.00	0.00	1.00
Master	640	0.09	0.29	0.00	0.00	0.00
Other Education	640	0.62	0.49	0.00	1.00	1.00
Sample Mainz	640	0.51	0.50	0.00	1.00	1.00

Notes: This table presents summary statistics for all personal characteristics. See Table 7 for detailed variable descriptions. Missing values for *Risk Averse*, *Perceived Financial Literacy*, and *Precision Financial Literacy* are imputed using scikit-learn's k-nearest neighbor imputation, which utilizes all other characteristics. Due to the sensitivity of K-Means Clustering to outliers, I replace outliers identified by the interquartile range method with the nearest non-outlier values for the clustering exercise in section 4.3. However, the non-adjusted values shown in this table are used for the prediction task in section 4.4.

Table 38: Summary statistics of portfolio characteristics

	count	mean	std	25%	50%	75%
Return 1 Month	640	-0.05	0.01	-0.06	-0.05	-0.04
Return 3 Months	640	-0.04	0.05	-0.06	-0.03	-0.00
Return 1 Year	640	0.04	0.11	-0.00	0.07	0.12
Return 3 Years	640	0.73	0.45	0.39	0.60	1.05
Dividend	640	1.24	0.33	0.99	1.23	1.45
Price Earnings Ratio	640	16.03	4.02	13.29	16.06	18.88
Volatility	640	0.22	0.03	0.20	0.22	0.24
Share Price	640	36.94	10.89	28.84	35.89	44.34
Equity	640	20640121576.80	11486643203.40	14225905920.00	18568330240.00	22707740672.00
# Employees	640	68997.74	30569.66	45954.60	65834.90	87261.95
Revenue	640	23352281749.20	11507954662.37	15140761600.00	22105698304.00	30015200256.00
Debt Equity Ratio	640	2.57	0.86	1.95	2.43	3.06
ESG Score	640	22.94	6.33	18.00	23.00	28.00
# Sustainable d.	640	0.61	0.49	0.00	1.00	1.00

Notes: This table presents summary statistics for all portfolio characteristics of the portfolios constructed by the participants in the baseline decision. See Table 7 for detailed variable descriptions. In the clustering task, *ESG Scores* are used, while *# Sustainable d.* is utilized in the prediction task. Given the sensitivity of K-Means Clustering to outliers, I replace outliers identified by the interquartile range method with the nearest non-outlier values for the clustering exercise. However, the non-adjusted values shown in this table are used for the prediction task.

Table 39: Distribution of share characteristics among the 20 available stocks

	Dividend	Price / Earnings	Volatility	Share Price	Equity	# Employees	Revenue	Debt / Equity	Return 1 Month	Return 3 Months	Return 1 Year	Return 3 Years	ESG Score
0.10	0.44	7.59	0.16	12.39	2272816980	9613.80	2996430000	1.01	-0.12	-0.28	-0.35	-0.16	-0.39
0.20	0.54	8.69	0.18	19.57	3842881056	15100.20	3787234112	1.06	-0.11	-0.18	-0.25	-0.05	5.44
0.30	0.66	9.16	0.19	23.24	4511763200	24820.00	7814755200	1.27	-0.09	-0.17	-0.19	0.12	11.31
0.40	0.77	10.32	0.21	23.89	6016660800	35780.00	8965152000	1.39	-0.08	-0.12	-0.15	0.25	15.78
0.50	1.18	12.77	0.24	24.72	11014160000	51150.00	12609766260	2.06	-0.04	-0.09	-0.07	0.35	20.21
0.60	1.27	14.66	0.26	29.82	17690384000	64459.60	20112744000	2.95	-0.04	-0.06	-0.03	0.45	23.34
0.70	1.51	17.21	0.29	36.46	23173332000	81860.00	24601416000	3.58	-0.03	-0.05	0.05	0.47	29.10
0.80	2.01	21.29	0.32	42.51	26683243200	104780.00	34285440000	4.58	-0.02	-0.03	0.07	0.59	31.20
0.90	2.23	27.16	0.34	68.50	34223700000	191560.00	76978524000	8.21	-0.02	0.06	0.12	1.09	36.39
# obs.	20	20	20	20	20	20	20	20	20	20	20	20	20

Notes: This table presents the distribution of share characteristics for the original 20 stocks. The first column lists the respective percentiles, and the subsequent columns provide the names and values of each stock characteristic.

Table 40: Average values for the resulting three clusters of K-Means Clustering

	Cluster 0	Cluster 1	Cluster 2	p-value (0 vs 1)	p-value (0 vs 2)	p-value (1 vs 2)
Age	25.09	23.25	24.74	0.00	0.29	0.00
Income	2.44	1.69	2.28	0.00	0.34	0.00
Possess Stocks	0.75	0.11	0.45	0.00	0.00	0.00
Possess Bonds	0.18	0.03	0.08	0.00	0.03	0.06
Possess Funds	0.75	0.15	0.45	0.00	0.00	0.00
Agrees with Organisation	5.85	5.95	5.80	0.21	0.92	0.28
ESG Knowledge	0.74	0.24	0.29	0.00	0.00	0.31
ESG Risk Lower	4.63	4.97	4.56	0.21	0.58	0.10
ESG Return Lower	3.48	3.93	3.98	0.03	0.09	0.72
Female	0.35	0.83	0.45	0.00	0.10	0.00
Has Siblings	0.44	0.52	0.47	0.08	0.66	0.41
Impact Altruism	1.78	2.01	1.56	0.28	0.39	0.11
Warm Glow Altruism	1.03	2.13	1.37	0.00	0.06	0.01
Risk Averse	0.47	0.48	0.46	1.00	0.32	0.28
# Financial Literacy	8.77	6.54	7.38	0.00	0.00	0.00
Perceived Financial Literacy	8.07	5.10	6.45	0.00	0.00	0.00
Precision Financial Literacy	0.44	0.30	0.36	0.00	0.00	0.00
Bachelor	0.40	0.23	0.31	0.00	0.18	0.08
Master	0.19	0.02	0.13	0.00	0.23	0.00
Other Education	0.41	0.75	0.55	0.00	0.03	0.00
Sample Mainz	0.27	0.65	0.46	0.00	0.00	0.00
Return 1 Month	-0.05	-0.05	-0.05	0.19	0.05	0.27
Return 3 Months	-0.02	-0.02	-0.12	0.20	0.00	0.00
Return 1 Year	0.08	0.06	-0.08	0.01	0.00	0.00
Return 3 Years	0.81	0.69	0.68	0.00	0.01	0.34
Dividend	1.23	1.28	1.14	0.08	0.03	0.00
Equity	18610173882.78	18088541679.41	24273430446.02	0.23	0.00	0.00

Table 40 (continued):

	Cluster 0	Cluster 1	Cluster 2	p-value (0 vs 1)	p-value (0 vs 2)	p-value (1 vs 2)
# Employees	61561.89	65398.27	97659.85	0.14	0.00	0.00
Revenue	20527537397.55	21930096924.85	34119136721.98	0.17	0.00	0.00
Price Earnings Ratio	16.33	16.79	12.30	0.15	0.00	0.00
Volatility	0.22	0.22	0.27	0.07	0.00	0.00
Share Price	38.99	36.87	32.14	0.01	0.00	0.00
Debt Equity Ratio	2.22	2.55	3.30	0.00	0.00	0.00
ESG Score	22.88	24.02	18.75	0.03	0.00	0.00
Number of Observations	196	355	89			

Notes: This table presents mean values for the resulting clusters of the K-Means Clustering with k=3 for all variables used as inputs for PCA. P-values are the results of a Mann-Whitney-U-Tests which assesses differences between the clusters.

Table 41: Performance metrics for different model specifications

Panel A: Random Forest using different features

	Metric	Personal Characteristics	Trading Desk Characteristics	All Characteristics
0	Accuracy	0.55	0.83	0.90
1	Precision	0.69	0.88	0.91
2	Recall	0.47	0.84	0.93
3	F1 Score	0.56	0.86	0.92
4	ROC AUC	0.59	0.90	0.96

Panel B: XGBoost using different features

	Metric	Personal Characteristics	Trading Desk Characteristics	All Characteristics
0	Accuracy	0.62	0.88	0.92
1	Precision	0.68	0.86	0.95
2	Recall	0.72	0.95	0.91
3	F1 Score	0.70	0.90	0.93
4	ROC AUC	0.58	0.92	0.97

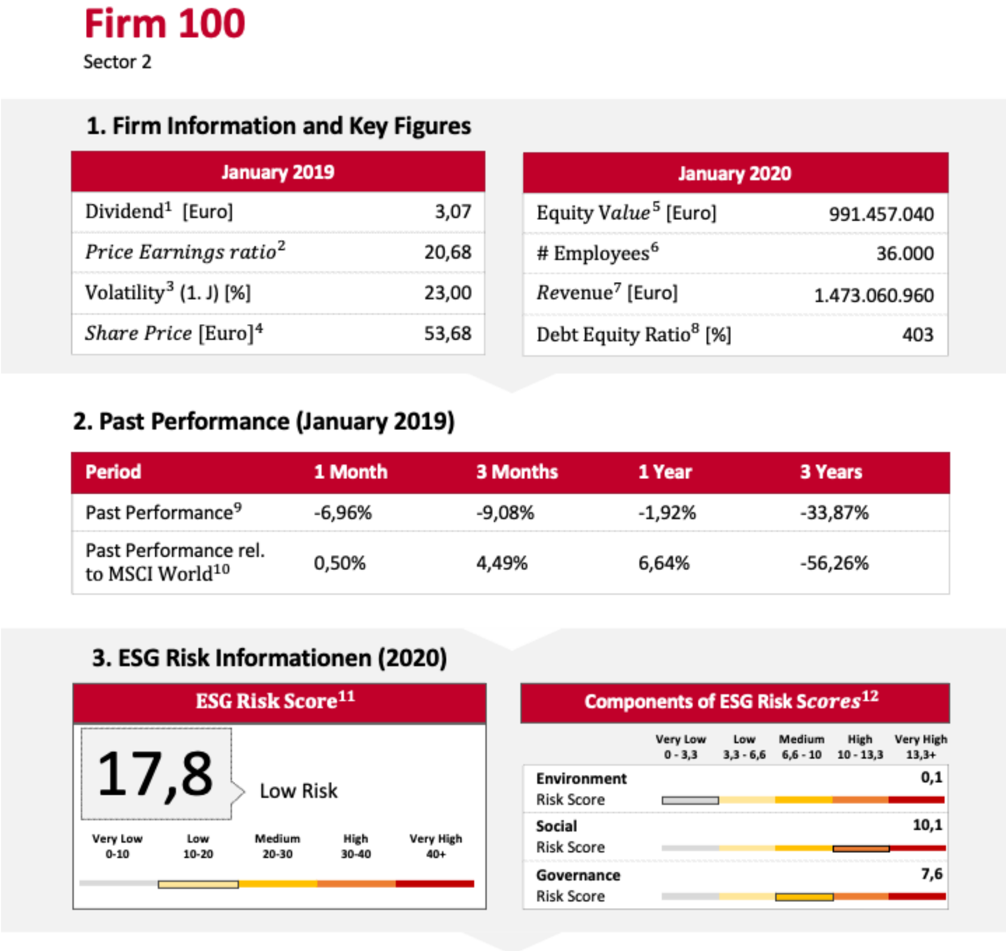
Panel C: K-Nearest-Neighbor using different features

	Metric	Personal Characteristics	Trading Desk Characteristics	All Characteristics
0	Accuracy	0.61	0.72	0.83
1	Precision	0.65	0.98	0.89
2	Recall	0.78	0.56	0.83
3	F1 Score	0.71	0.71	0.86
4	ROC AUC	0.55	0.85	0.91

Panel D: Logistic Regression using different features

	Metric	Personal Characteristics	Trading Desk Characteristics	All Characteristics
0	Accuracy	0.65	0.80	0.90
1	Precision	0.64	0.83	0.94
2	Recall	0.96	0.85	0.89
3	F1 Score	0.77	0.84	0.91
4	ROC AUC	0.53	0.86	0.95

Notes: This table displays the performance metrics for different machine learning models based on three different sets of features: (i) personal characteristics alone, (ii) personal characteristics combined with portfolio characteristics visible on the trading desk, and (iii) all available personal and portfolio characteristics. Panel A shows the result of the Random Forest model, Panel B of the XGBoost model, Panel C of the K-Nearest-Neighbor model, and Panel D of the Logistic Regression model.



Firm 100

Sector 2

1. Firm Information and Key Figures

January 2019	
Dividend[1] [Euro]	3,07
Price Earnings ratio[2]	20,68
Volatility[3] (1. J) [%]	23,00
Share Price [Euro][4]	53,68

January 2020	
Equity *Value*[5] [Euro]	991.457.040
# Employees[6]	36.000
Revenue[7] [Euro]	1.473.060.960
Debt Equity Ratio[8] [%]	403

2. Past Performance (January 2019)

Period	1 Month	3 Months	1 Year	3 Years
Past Performance[9]	-6,96%	-9,08%	-1,92%	-33,87%
Past Performance rel. to MSCI World[10]	0,50%	4,49%	6,64%	-56,26%

3. ESG Risk Informationen (2020)

ESG Risk Score[11]

17,8 Low Risk

Very Low	Low	Medium	High	Very High
0-10	10-20	20-30	30-40	40+

Components of ESG Risk *Scores*[12]

	Very Low 0 - 3,3	Low 3,3 - 6,6	Medium 6,6 - 10	High 10 - 13,3	Very High 13,3+	
Environment Risk Score						0,1
Social Risk Score						10,1
Governance Risk Score						7,6

Figure 17: Fact sheet of stock 100

Notes: This Figure displays a typical fact sheet of a stock in the stock market game of the experiment. By clicking on the stock's name in the trading desk, subjects obtained this kind of information. It provides information about a stock from 2019, including the dividend, PE-ratio, Volatility, Share Price and Past Performance. Additionally, it contains information from 2020, namely the Equity value, number of Employees, Revenue, Debt Ratio, and some more details about the ESG Risk Score.

Figure 18: Characteristics of the three personas derived from Clustering with $k = 3$

Notes: This figure displays the personas derived from the clustering analysis with $k = 3$ (refer to Table 40). The persona on the left depicts Persona C, a young male sustainable investor from Frankfurt. The persona in the middle describes persona D, a young male conventional investor also from Frankfurt. Lastly, the persona on the right demonstrates persona A, a young sustainable investor affiliated with Mainz.

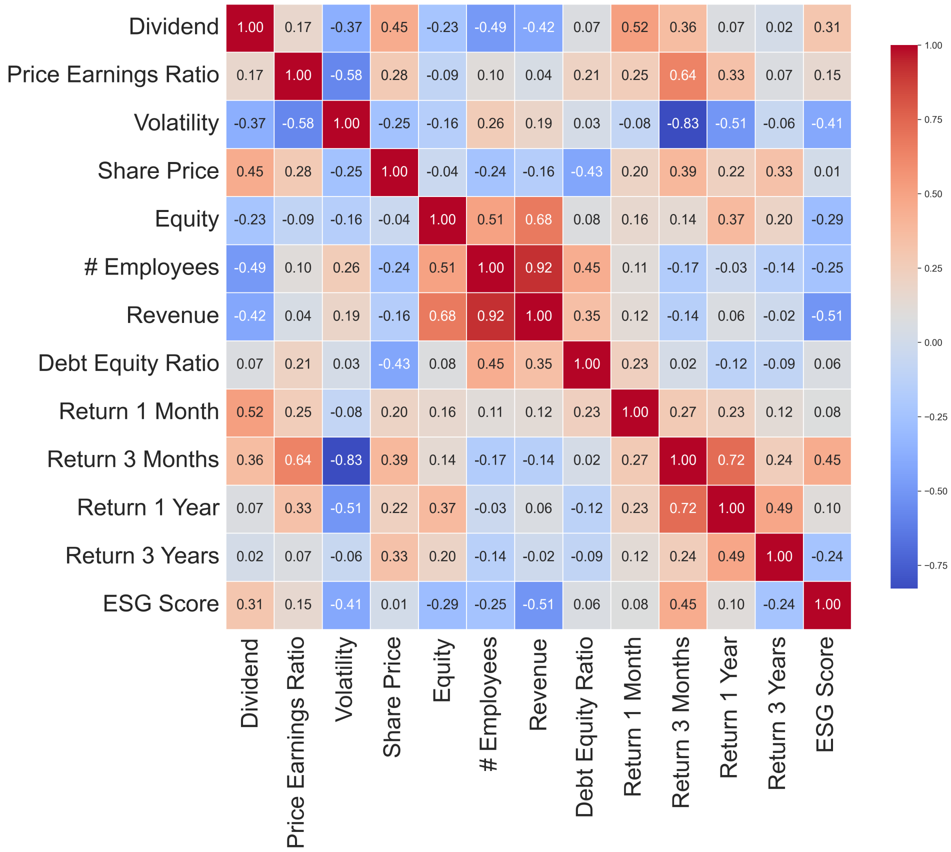


Figure 19: Correlation portfolio characteristics (original stock universe)

Notes: This Figure displays the Spearman rank correlation of the portfolio characteristics among the 20 stocks from the original stock universe. It highlights a strong negative correlation between *volatility* (*revenue*) and ESG Scores, alongside a strong positive correlation between *Return 3 Months* and ESG Scores.

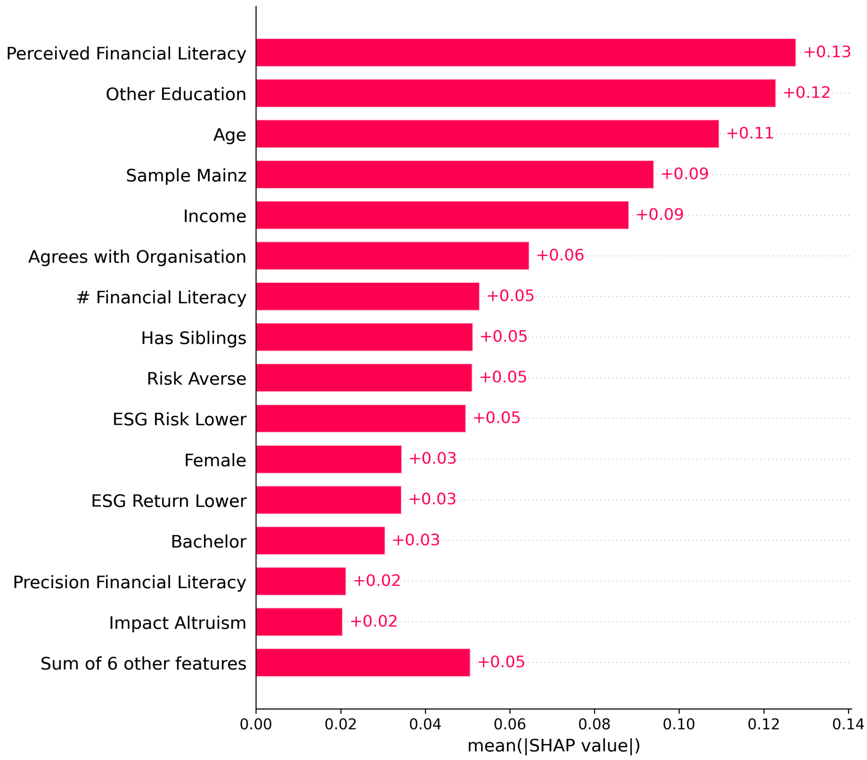


Figure 20: Mean SHAP Plot (Random Forest model using personal characteristics)

Notes: This figure illustrates the SHAP feature importance of the Random Forest Model using personal characteristics by averaging the absolute Shapley values per feature across the dataset. *Other Education, Age, Perceived Financial Literacy, Agrees with Organization, Income*, and *Bachelor* emerge as the most significant features.

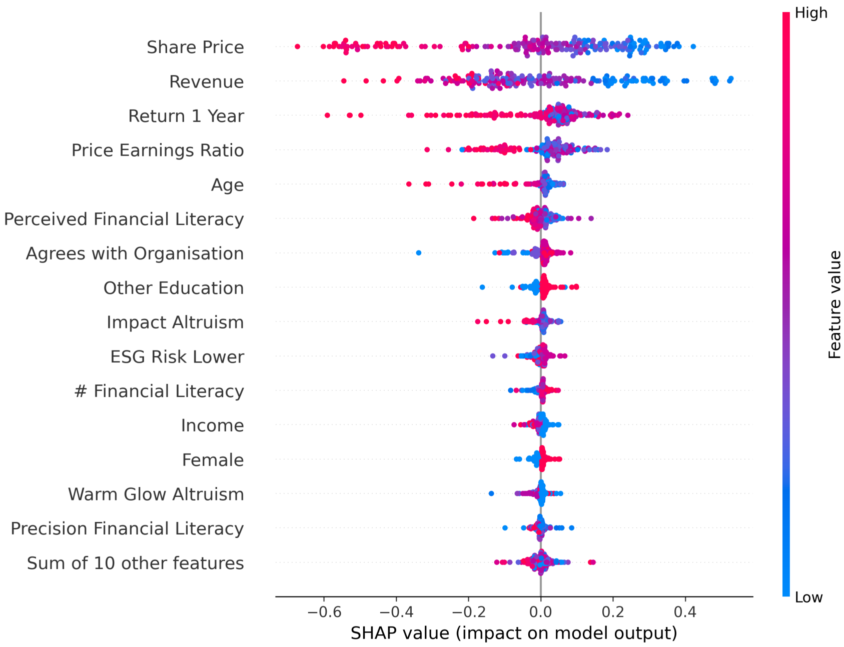


Figure 21: SHAP Beeswarm Plot (Random Forest model using personal and most salient portfolio characteristics)

Notes: Figure 21 shows that the four portfolio characteristics are the most influential in explaining the predictions, with *Share Price* being the most impactful. There exists a negative correlation between being a sustainable investor and both *Share Price* and *Revenue*, and a positive correlation between being a sustainable investor and both *Return 1 Year* and *Price Earnings Ratio*.

Chapter 5

5 What Motivates Wealthy Private Investors to Hold Sustainable Investments?

Fabian Braun[ab] Andrej Gill[ac] Florian Hett[ad]

August 2024

Abstract

What influences sustainable investment behavior of wealthy private investors? We combine data from an incentivized online experiment with the Investor Suitability Assessment of a German private bank to study the role of altruistic preferences and risk and return expectations in sustainable investment decisions. Our primary findings are threefold: (i) We detect a significant correlation between altruistic preferences and sustainable investments, predominantly driven by impact altruism. (ii) Investors are willing to spend on average 5.6% of their budget on ESG Score information that facilitates sustainable investment decisions. (iii) Exposure to information indicating overperformance (underperformance) of sustainable stocks leads to an increase (decrease) in ESG Scores relative to the baseline portfolio allocations of 22.8% (12.2%). Our results have implications for policymakers aiming to attract private funds to finance the green transition of the economy, for regulators defining guidelines for sustainable investments, and for financial institutions providing such products.

Keywords: ESG, Sustainable Finance, Wealthy Private Investors, Stock Market Game, Experiment

[a]Comments from audiences at Goethe University Frankfurt, Leibniz Institute for Financial Research SAFE, and Johannes Gutenberg University Mainz are gratefully acknowledged. The joint ethics committee of the Faculty of Economics and Business Administration of the Goethe University Frankfurt and the Gutenberg School of Management & Economics of the Faculty of Law, Management and Economics of the Johannes Gutenberg University Mainz granted the project exempt status. The experiment is pre-registered under AEARCTR-0009341 and, unless otherwise stated, we are following a pre-registered analysis plan. We thank Lea Voigt for excellent research assistance. The authors declare no conflict of interest.

[b]Johannes Gutenberg University Mainz; Jakob-Welder-Weg 9, 55128 Mainz, Germany. Phone: +49 6131 3927901. Email: braun@uni-mainz.de

[c]Johannes Gutenberg University Mainz; Jakob-Welder-Weg 9, 55128 Mainz, Germany. Phone: +49 6131 3922979. E-Mail: gill@uni-mainz.de

[d]Johannes Gutenberg University Mainz; Johann-Joachim-Becher-Weg 31, 55128 Mainz, Germany. Phone: +49 6131 39-20567. E-Mail: florian.hett@uni-mainz.de

5.1 Introduction

Financing the green transition to achieve net zero emissions by 2050 requires an estimated 7 trillion USD annually from both private and public capital. However, current investments fall short, with less than 2 trillion USD being invested annually (Cartry *et al.*, 2023). According to the Global Wealth Report by Credit Suisse (2023), individuals with net wealth between 100,000 USD and 1 million USD own net assets totaling 178.9 trillion USD, representing 39.4% of privately owned financial wealth.[71] Given their substantial financial resources, these individuals are a crucial target group for funding the green transition. Yet, most existing research in sustainable finance focuses on demographics with lower wealth levels. These studies demonstrate a strong preference for sustainable investments among investors (Barreda-Tarrazona *et al.*, 2011; Hartzmark and Sussman, 2019; Bauer *et al.*, 2021), driven by key factors such as altruistic preferences (Riedl and Smeets, 2017; Bauer *et al.*, 2021; Heeb *et al.*, 2023) and risk and return expectations (Døskeland and Pedersen, 2016; Braun *et al.*, 2024b). This paper seeks to extend these findings by showing distinctive characteristics of experienced wealthy private investors interested in sustainable investments.

We collaborate with a German private bank that serves wealthy private investors, enabling us to conduct an online experiment with their clientele. We introduce an online stock market game designed to elicit revealed preferences for socially responsible investments (SRI). In this game, participants allocate money across real stocks, whose names are masked to prevent the use of external information. Our experimental setup introduces a controlled information treatment about the performance of sustainable stocks and observes investors' reactions within the game. To ensure incentive compatibility, we pay out one of the constructed portfolios to the investors after the experiment and make actual stock market investments based on the choices made during the game. Additionally, the study incorporates incentive-compatible measures of individual altruism (Andreoni, 1989, 1990), survey measures of financial literacy (Hastings *et al.*, 2013; Lusardi and Mitchell, 2008), risk preferences (Falk *et al.*, 2018), expectations about the risks and returns of SRI (Riedl and Smeets, 2017), ESG knowledge, and demographic data. Finally, the experimental data is combined with administrative data from the bank's Investor Suitability Assessment[72] which includes information on income, wealth, risk preferences, experience with various financial products, and the age and education of the participants.

The primary findings of this study can be summarized in three key areas: (i) We establish a preference for sustainable investments among wealthy private investors and detect a significant correlation between altruistic preferences and sustainable investments. For investors lacking prior ESG knowledge, this correlation is predominantly driven by impact altruism, whereas for

[71] 13% of the global population possesses wealth exceeding 100,000 USD. This affluent demographic controls over 387 trillion USD, which represents 85% of privately owned financial wealth (Credit Suisse, 2023).

[72] MiFID II requires institutions that provide financial advice or portfolio management to do an Investor Suitability Assessment, where investors' experience, financial situation, and investment objectives are elicited (ESMA, 2022).

those with ESG knowledge, both impact and warm glow-driven altruistic preferences play a role. (ii) Our results show that investors are willing to spend, on average, 5.6% of their budget to obtain ESG Score information, underscoring a significant willingness to pay for information that facilitates sustainable investment decisions. (iii) We observe that investment behavior is highly responsive to information about the risk-adjusted performance of sustainable investments. Exposure to information indicating overperformance (underperformance) of sustainable stocks leads to an increase (decrease) in ESG Scores relative to baseline portfolio allocations of 22.8% (12.2%). Collectively, these findings corroborate those reported by Braun *et al.* (2024b) for a sample of younger demographics, demonstrating consistency across diverse investor profiles.

Overall, our study extends the findings of the sustainable finance literature to a distinct demographic: wealthy private investors with extensive investment experience. Instead of relying solely on survey evidence or hypothetical scenarios (e.g., Paetzold and Busch (2014); Gutsche and Ziegler (2019)), we integrate data from the bank's Investor Suitability Assessment, which includes information on income, wealth, risk preferences, experience with financial products, age, and education. This data shows strong consistency with investors' self-reported information. Additionally, we elicit revealed sustainability preferences in an incentivized stock market game, enabling us to address concerns about experimenter demand effects (Haaland *et al.*, 2023) and hypothetical bias (List and Shogren, 1998; List, 2001; Harrison, 2006).

Our findings corroborate in an experimental setting that wealthy private investors overall value sustainability (Barreda-Tarrazona *et al.*, 2011; Paetzold and Busch, 2014; Hartzmark and Sussman, 2019; Bauer *et al.*, 2021; Braun *et al.*, 2024b). Key studies underscore the significance of altruistic preferences in driving sustainable investment decisions (Riedl and Smeets, 2017; Bauer *et al.*, 2021; Heeb *et al.*, 2023; Gutsche *et al.*, 2023). For example, the seminal work by Riedl and Smeets (2017) integrates administrative, survey, and experimental data from Dutch investors, demonstrating that social preferences, rather than financial gains, predominantly motivate sustainable investment choices. In addition, the literature debates whether sustainable investments are motivated by warm glow or impact altruism[73] (Riedl and Smeets, 2017; Heeb *et al.*, 2023; Gutsche *et al.*, 2023). This distinction is essential for establishing regulatory frameworks that support true sustainable investments and for developing investment strategies that align with investor motivations. Heeb *et al.* (2023) argue that although investors show a substantial willingness to pay for sustainable investments, their payments do not scale with the potential impact, suggesting that warm glow rather than impact altruism drives their investment choices. Prior research that experimentally elicits altruistic preferences concentrates on either impact or

[73] According to Andreoni (1989, 1990), the difference between these concepts is that warm glow altruism represents a selfish motivation for altruistic behavior, where individuals do good deeds to feel good about themselves. Conversely, impact altruism involves altruistic behavior without expecting any personal reward, motivated solely by the benefits to others. See Andreoni (1989, 1990) for a discussion on the different underlying motivations behind charitable giving. He introduces the concepts of pure altruism and warm glow, which we refer to as impact altruism and warm glow altruism.

warm glow altruism and therefore neglects the interaction of these variables. Alongside Braun *et al.* (2024b), to our knowledge this is the first study to examine the simultaneous correlation of warm glow and impact altruism on sustainable investments. Our research extends the literature by establishing a significant correlation between experimentally elicited measures of impact altruism and sustainable investments controlling for warm glow altruism among wealthy private investors.

In addition, we contribute to the literature on the effects of risk and return expectations on sustainable investments (Døskeland and Pedersen, 2016; Riedl and Smeets, 2017; Hartzmark and Sussman, 2019; Dong *et al.*, 2022; Gutsche *et al.*, 2023; Giglio *et al.*, 2023; Braun *et al.*, 2024b).[74] Giglio *et al.* (2023) survey Vanguard clients and observe that, although investors generally expect sustainable investments to underperform the market, only those anticipating higher returns hold significant ESG assets. Conversely, Riedl and Smeets (2017) incorporate risk and return expectations alongside social preferences in their regression framework and conclude that non-pecuniary motives drive sustainable investments. Our study is distinguished by the exogenous variation in risk and return expectations among participants and the subsequent observation of their behavior in an incentivized stock market game. This approach avoids reliance on potentially overly optimistic survey items on risk and return expectations (Dong *et al.*, 2022) or outcome variables subject to hypothetical bias (List and Shogren, 1998; List, 2001; Harrison, 2006; Harrison *et al.*, 2008). Most similar to our methodology is Døskeland and Pedersen (2016), who, in a natural field experiment within an online banking context, demonstrate that financial considerations significantly motivate investors to select sustainable stocks. Unlike their study, which employs positive wealth framing, we also investigate the effects of negative performance information on sustainable assets and demonstrate that this reduces sustainable investment behaviors. More broadly, we contribute to debates on whether individuals driven by social reasons might decrease their intrinsic motivation to invest sustainably if their actions are perceived as financially attractive (Frey and Oberholzer-Gee, 1997; Mellström and Johannesson, 2008; Gneezy *et al.*, 2011; Brodback *et al.*, 2019). Our findings elucidate that the investment behavior of wealthy private investors significantly responds to changes in both, increases and decreases in risk and return expectations concerning sustainable investments. This is consistent with an explanation that increases (decreases) in financial utility outweigh decreases (increases) in social utility.

Last, we add to the literature on investors' willingness to pay to invest sustainably (Gutsche and Ziegler, 2019; Heeb *et al.*, 2023; Baker *et al.*, 2022; Laudi *et al.*, 2023; Engler *et al.*, 2023).

[74] Results on the performance of sustainable stocks are mixed. While some studies argue that investors might expect higher returns for non-sustainable assets due to theoretical and empirical evidence (Hong and Kacperczyk, 2009; Chava, 2014; Pedersen *et al.*, 2021; Bolton and Kacperczyk, 2021,?; Pastor *et al.*, 2021; Avramov *et al.*, 2022; Bolton and Kacperczyk, 2024), others suggest that the performance of sustainable firms is not worse, or even better in the long run (Kempf and Osthoff, 2007; Eccles *et al.*, 2014; Dimson *et al.*, 2015; Friede *et al.*, 2015; Avramov *et al.*, 2022; Berg *et al.*, 2022).

Baker *et al.* (2022) show that on average investors are willing to pay 20 basis points higher fees for ESG-oriented index funds as compared to otherwise identical funds without an ESG mandate. According to Laudi *et al.* (2023) financial advisor use this higher willingness to pay and especially charge higher fees for financially illiterate individuals. This premium is not justified by the effort, skill, or costs involved. We add to that discussion by showing that investors are willing to pay for ESG Scores and that financially illiterate investors are willing to pay significantly more for ESG Scores than financially literate investors. Laudi *et al.* (2023) discuss that this form of price discrimination could be problematic by potentially harming consumer welfare and diminishing the long-term appeal of sustainable investments. We differ from the existing literature in a way, that we directly measure investors' willingness to pay for ESG Scores, thus giving implications to policymakers not only to analyze higher fees for sustainable products but also keep track of the pricing of those information sources helping to make sustainable investment decisions.

The remainder of the paper is structured as follows: In section 5.2 we provide details on the experimental design, the data collection, and key variables. Section 5.3 gives information on the institutional background of the partnering institution, the administrative data obtained by the partnering bank and an overview on sample characteristics. Section 5.4 presents the main results. Section 5.5 discusses the results and concludes.

5.2 Experimental Design, Data Collection, and Key Variables

In this paper, we integrate revealed preferences elicited through incentive-compatible experiments with survey measures, demographics, and administrative data from our partnering bank. Our experimental design mirrors Braun *et al.* (2024b), which focus on a student sample with less experienced investors. Utilizing the same method for eliciting sustainability preferences as this prior study allows us to compare our results and assess the generalizability of the findings across different investor profiles.[75]

Measurement of Revealed Sustainability Preferences: In order to elicit investors' revealed sustainability preferences, we utilize an incentivized stock market game, allowing for precise control over the information available to participants and emphasizing actual decision-making. This approach not only minimizes hypothetical bias (List and Shogren, 1998; List, 2001; Harrison, 2006; Harrison *et al.*, 2008) but also reduces the potential for experimenter demand effects (Haaland *et al.*, 2023), as participants make real financial choices with tangible outcomes. Each participant is equipped with a trading account to buy up to five stocks from a pool of 20, providing a representative cross-section of the market.[76] In the experiment, we use real share

[75]See our pre-analysis plan (AEARCTR-0009341) for the exact wording of the experimental instructions. We mainly follow stages 1 to 3 of Braun *et al.* (2024b) with the minor difference that, to prevent dropouts, we had to shorten the duration of the experiment and therefore did not incentivize the elicitation of the willingness to pay in Stage 2.

[76]The selection involved 100 listed companies from Yahoo Finance, filtered by several criteria: 1) Actual companies, excluding funds; 2) Fiscal year ending in December; 3) Classification into ESG Risk Score ranges of

prices and attributes but anonymize the names and substitute prices with an experimental currency (points). Any market gains or losses translate directly to the experimental currency, thus mirroring real-world stock behavior within our controlled environment. Figure 22 provides a screenshot of the trading interface used in the game. When participants hover their cursor over one of the key figures, a tooltip offers a brief explanation of the indicator. Additionally, clicking on the stock name opens a detailed fact sheet presenting ten key figures for that stock, enhancing transparency and aiding decision-making (see Figure 23 in the appendix for an example).

Account Balance:
50.000 Punkte

Name	Share Price (2019)	Past Performance (2018-2019)	P/E Ratio (2019)	Revenue (2020)	ESG Risk Scores (2020)
Firma 56	100,76	-17,56%	11,55	45.192.000.000	39,3
Firma 80	37,28	6,77%	29,39	31.303.440.000	25,1
Firma 7	12,92	4,78%	28,08	80.185.000.000	16,8
Firma 27	19,71	-15,00%	8,89	76.644.960.000	28,0
Firma 76	36,10	0,09%	8,06	3.160.200.000	7,8
Firma 94	19,02	6,07%	14,07	8.585.892.000	16,0

Figure 22: Trading desk

Notes: This figure presents the (translated) trading desk interface used by participants during the stock market game. Each participant is endowed with 50,000 points and can allocate this endowment among five different stocks selected from an assortment of 20. Initially, participants view key figures for each stock such as Share Price, Past Performance, Price Earnings Ratio, revenue, and ESG Risk Score. By clicking on the stock's name, participants receive additional information, as illustrated in Figure 23, and have the option to purchase the stock.

Participants are provided with publicly available information on past performance, dividends, price-to-earnings ratios, volatility, and the stock price as of January 2019. Furthermore, they are presented with data from 2020 concerning equity value, number of employees, revenue, debt ratio, and ESG Risk Score. Control questions are introduced to ensure participants can differentiate between the data from 2019 and 2020. Each participant commences the trading process with a trading account containing 50,000 points, which may be utilized to purchase up to five different stocks at January 2019 prices from a pool of 20 stocks. This is accomplished by clicking the "Buy" button positioned above and below the fact sheet.[77] The experiment is conducted in two rounds. Initially, we elicit baseline demand for sustainable investments. After an information treatment regarding the risk-adjusted performance of sustainable stocks, participants enter the game a second time. This approach yields three distinct measures of sustainability for each portfolio allocation: The first, *ESG Score*, is the average ESG Score of the selected five stocks.[78] Higher

0-10, 10-20, 20-30, 30-40, 40+; 4) Market cap categorizations as small/mid-cap, large-cap, or mega-cap. The selection of the final 20 companies was random from each bin and ensured proportional representation from each ESG category and market cap level. Due to a lack of sustainable firms in the largest market cap bin, firms from the next smaller bin were used.

[77] Participants are required to select exactly five stocks, but the same stock can be selected multiple times.

[78] In the experiment, participants view the unprocessed ESG Risk Scores from Sustainalytics, where a higher

ESG Scores imply greater sustainability. As this indicator provides the richest information on sustainability, it is preferred throughout the paper. The second measure, *# Sustainable*, counts the number of sustainable stocks in a portfolio.[79] The third measure, *# Sustainable d.*, is an indicator variable set to one if the portfolio includes more than two sustainable stocks.[80]

Measurement of Preferences, Expectations, and Demographics: The experimental literature identifies several factors correlated with socially responsible investments, including altruism, risk aversion, risk and return expectations, financial literacy, gender, age, income, and wealth (Riedl and Smeets, 2017; Gutsche and Ziegler, 2019; Bauer *et al.*, 2021; Haber *et al.*, 2022; Gutsche *et al.*, 2023; Montagnoli and Taylor, 2024). To elicit incentive-compatible measures of altruism, we utilize two dictator games paired with a charity where participants allocate ten Euros between themselves and a selected charity (Crumpler and Grossman, 2008; Tonin and Vlassopoulos, 2010). The experimental design incorporates two types of dictator games: a standard dictator game and a "money-burning" dictator game. Unlike the standard game, in the "money-burning" game, the charity receives ten Euros regardless of the participant's donation. This setup facilitates the measurement of three distinct aspects of altruism: (i) *altruism*, reflecting a combination of warm glow and impact altruistic motivations, measured by the donation amount in the standard game; (ii) *warm glow altruism*, derived from the donation in the "money-burning" game, signifying the satisfaction from the act of giving—hence, giving by others does not act as a perfect substitute (Andreoni, 1989, 1990); (iii) *impact altruism*, calculated as the difference between donations in the two games, indicating the utility derived from the benefit of the recipient of the donation, with impact altruists considering the giving of others as a perfect substitute (Andreoni, 1989, 1990). Moreover, we elicit risk preferences (Falk *et al.*, 2018), risk and return expectations (Riedl and Smeets, 2017), financial literacy (Hastings *et al.*, 2013; Lusardi and Mitchell, 2008), ESG knowledge, the future importance of ESG and a set of demographics to get a profound understanding of investors knowledge and beliefs. Table 47 in the appendix provides a detailed definition of each of the variables.

Measurement of Willingness to Pay (*WTP*) for ESG Score information: The *WTP* for ESG Scores holds significant interest for both financial institutions, which aim to develop new products, and regulators, who seek to understand individual preferences for sustainability information for regulatory purposes. To measure investors' *WTP* for ESG Score information, participants are presented with a hypothetical scenario where, unlike in the previously explained stock market game, the ESG Score information is not provided for free. They are asked to express their *WTP* using a slider ranging from 0 to 10,000 points.

score indicates a lower level of sustainability. For the sake of clarity, the ESG Risk Scores are transformed into ESG Scores using the following formula: *ESG Score = (2*mean(ESG Risk Score) - ESG Risk Score).*

[79] Sustainable stocks are defined by an ESG Risk Score of less than 20 (color-coded in grey and yellow).

[80] A random selection would typically yield two sustainable stocks.

Reaction to Performance Information of Sustainable Stocks: There is an ongoing academic debate about whether incorporating ESG criteria into investment processes increases or decreases risk-adjusted returns (see Hong and Kacperczyk (2009); Eccles *et al.* (2014); Friede *et al.* (2015); Pedersen *et al.* (2021); Bolton and Kacperczyk (2021); Pastor *et al.* (2021); Briere and Ramelli (2021); Avramov *et al.* (2022); Bolton and Kacperczyk (2023); Latino (2023); Bolton and Kacperczyk (2024) among others). To investigate investors' causal reactions to changes in return expectations, we exogenously vary the information provided about the risk-adjusted performance of firms with high (low) sustainability ratings. Specifically, we randomly assign 50% of participants to receive a management summary titled *"Why investing in stocks with low ESG Risk Scores makes sense from a performance perspective and from a scientific point of view"*, while the other half receive a summary titled *"Why investing in stocks with low ESG Risk Scores does not make sense from a performance perspective and from a scientific point of view"*. After reviewing these summaries, participants engage in the stock market game described previously for a second time.

Data Collection, Payment of Participants, and Incentive Compatibility: Data collection was conducted using an oTree-programmed online experiment (Chen *et al.*, 2016), targeting wealthy private investors from a German private bank between May 10th and September 20th, 2022. We contacted 2,000 clients via the bank's online banking system, achieving a response rate of 2.23%, which aligns with similar studies (Giglio *et al.*, 2021; Famulok *et al.*, 2023). The final sample consists of 74 participants.[81] The stock market game is incentivized; each participant receives a payout based on the value of one of the two portfolios as of January 2020, using an exchange rate of 1 point = 0.0003 Euros. Participants are also informed that the randomly selected portfolio will be purchased on the real stock market, with 20% of its actual value invested using private funds. This ensures that investment decisions during the experiment have direct financial consequences, thereby imposing moral responsibility on the participants. Each portfolio is maintained for 12 months. Additionally, one decision from either the standard dictator game or the "money-burning" dictator game is paid out to participants (and the corresponding charities). All payments to participants and charities were made after the experiment's conclusion. The median time spent in the experiment was 37 minutes, and average earnings were 22 Euros per participant, equivalent to an hourly wage of 36 Euros.

5.3 Institutional Background and Sample Characteristics

The market for sustainable investments is full of conflicting information (Braun *et al.*, 2024b). The European Union identified these information frictions as challenging for investors want-

[81] We initially pre-registered 140 participants. To boost participation, we sent a reminder, resulting in an additional 16 participants that day; however, subsequent participation declined. Due to restrictions from our partnering bank, further reminders were not permissible. The experiment was terminated on September 20th after more than 60 days without any new participant activity.

ing to buy sustainable investments and introduced four key legislative initiatives to increase transparency and facilitate the effective transmission of sustainability preferences.

Key Sustainability Regulations: In the past, there existed only a limited common understanding of the reporting of sustainability information for firms, making it challenging to compare the available information between companies. The Corporate Sustainability Reporting Directive (CSRD) (Commission, 2022b) addresses this issue for large European companies as well as non-European companies that generate over 150 million Euros in the EU market. Starting with the fiscal year 2024, companies must publish regular reports on the environmental and social impact of their activities (Commission, 2024).[82]

ESG ratings are often used to identify the sustainability of firms. However, different ESG rating agencies not only use different inputs but also differ in their methodologies, resulting in highly heterogeneous sustainability ratings for the exact same firms across different rating agencies (Dimson *et al.*, 2020; Billio *et al.*, 2021; Avramov *et al.*, 2022; Berg *et al.*, 2022). The EU Taxonomy (Commission, 2020b) addresses the issue that identifying sustainable companies can be quite challenging and intends to establish a common understanding of sustainable economic activities. The regulation entered into force on July 12, 2020, and classifies economic activities as sustainable that contribute to at least one of the European Union's climate and environmental objectives.[83] At the same time, these activities must maintain minimum social safeguards and cannot harm any other environmental objective. Financial companies can use these definitions to define sustainable products.

The Sustainable Finance Disclosure Regulation (SFDR) (Commission, 2019b) complements the CSRD and the EU Taxonomy by creating a detailed reporting framework for financial products and entities and came into effect on March 10, 2021. The regulation defines an investment as sustainable if (i) it contributes to an environmental (E) or social (S) objective, (ii) does not significantly harm any other environmental or social objective, and (iii) the invested company follows good governance practices (G). Furthermore, it defines the principal adverse impacts (PAI) of economic activities, a list of 47 indicators that measure sustainability risks that can be used to define the sustainability of an investment.[84] Although the SFDR requirements are linked to the EU Taxonomy by including the specified sustainable activities, currently there exists a lack of consistency between the two regulations (Cordes *et al.*, 2024).

[82]The CSRD will also apply to listed SMEs for fiscal years starting 2026 (Commission, 2024). In addition, the IFRS foundation developed International Sustainability Standards, which intend to define global minimum standards in sustainability disclosure (IFRS, 2024) and apply from the fiscal year 2024 onwards but are not mandatory.

[83](i) climate change mitigation; (ii) climate change adaptation; (iii) water protection; (iv) contribution to a circular economy; (v) pollution prevention; (vi) protection of biodiversity and ecosystems (Commission, 2021b).

[84]The SFDR contains 14 mandatory and 33 voluntary PAIs. The 14 mandatory PAIs include greenhouse gas emissions, carbon footprint, and carbon intensity among others.

The amendment of the Markets in Financial Instruments Directive II (MiFID II) (Commission, 2021a) is designed to ensure the suitability of a financial product for an investor and to prevent the dissemination of misleading information regarding sustainable investment products (ESMA, 2023). Therefore, investor protection mandates that institutions providing financial advice or portfolio management must conduct an Investor Suitability Assessment before recommending or selling investment products (ESMA, 2022). This requires the elicitation of clients' knowledge and experience, financial situation, and investment objectives. Such information is typically gathered using (digital) questionnaires completed by the clients or collected during discussions with them (ESMA, 2023). The final guidelines on the MIFID II suitability requirements came into force in October 2023 and additionally aim to incorporate clients' sustainability preferences (ESMA, 2023). Specifically, this entails inquiring of clients whether they desire to include financial products that: (i) allocate a specified minimum percentage to green investments as defined by the EU Taxonomy Regulation; (ii) dedicate a minimum percentage to sustainable investments according to the SFDR; or (iii) consider the PAIs of an investment. Additionally, clients specify the minimum proportion of sustainable products in their overall investment. The objective of this process is to provide clients with comprehensive control over their sustainability preferences, offering a spectrum of potential sustainability definitions.

Administrative Data: To comply with MiFID II, the investment advisors of our partnering bank elicit the personal and financial situation of their clients by discussing a series of questions on monthly net income, existing wealth, risk tolerance, financial experience (knowledge), investment objective, and ESG preference, and then completing (updating) a predefined questionnaire. To enrich our analysis, we integrate the bank's data of this Investor Suitability Assessment with our experimental and survey data. The administrative data encompasses information on investors' risk attitudes and experiences. The variable *risk preferences (financial instrument)* captures investors' responses to the question *"What level of risk are you prepared to take as part of your investment strategy for individual securities investments?"*, rated on a scale from one to five. Investors also answer the question *"What experience and knowledge do you have in various categories of securities?"*, which assesses their familiarity with different investment products and provides insights into the average number and amount of their annual trades. The experience categories are ranked according to the increasing risk level of the underlying assets. Here, *experience riskiest asset class* is an indicator variable set to one for investors experienced in the riskiest asset classes (options, cum certificates, knock-out certificates, or hedge funds), and zero otherwise. Additionally, the bank supplies key demographic details such as *age, education, income, expenses*, and *wealth* of the participants in the experiment. As the regulation on the elicitation of sustainability preferences came into effect in October 2023 (ESMA, 2023), the bank has not updated that information on all their clients and therefore could not provide us with data on this item.[85] Table 47 (appendix) provides exact definitions for each variable.

[85]The bank collects this information if the investment advisor has contact with the client. As there are no

Data quality and consistency: The following analyses are exploratory. As a consistency check, we compare the investors' self-reported age in the experiment with the administrative data obtained from the bank. All participants accurately disclosed their age, which instills confidence in the veracity of responses to our survey questions. Additionally, we explore the correlation between the global preference survey risk item (*risk preferences*) and the risk and experience measures of the bank´s Investor Suitability Assessment (*risk preferences (financial instruments)* and *experience riskiest asset class*) using an ordinary least squares (OLS) specification (see Table 48 in the appendix). In Column (1) of the Table, we demonstrate that a one standard deviation increase in the maximum willingness to take risks for a single financial asset is associated with a 0.39 standard deviation increase in the global preference survey risk measure ($p < 0.01$). Moreover, individuals with knowledge and experience in options, cum certificates, knock-out certificates, or hedge funds exhibit a 0.73 standard deviation higher willingness to take risks ($p < 0.01$). Specification (3) confirms these findings and indicates that the two risk and experience measures elicited by administrative data of the bank are complements rather than substitutes, as evidenced by the increase in the adjusted R^2 value.

Unfortunately, institutional reasons prevented the bank from providing us with true sustainability portfolio holdings for each investor. Consequently, we use the two portfolio allocations in the stock market game as a revealed sustainability measure. Table 49 (appendix) presents OLS specifications demonstrating significant correlations between the mean *ESG Score* from the first and second portfolio allocations and various stated sustainability measures. Specification (1) shows a statistically significant correlation between the indicator variable *plan SRI* and the *ESG Score* in the baseline portfolio allocation ($p = 0.06$), while *possess SRI* shows a positive but not statistically significant correlation ($p = 0.24$). Individuals planning to invest sustainably within the next three years have portfolios with *ESG Scores* that are 19% higher compared to those who are either uninterested in or unsure about their current ownership of sustainable investments. Specification (2) combines individuals who plan to invest and those who already possess sustainable investments, showing a positive but marginally non-significant coefficient ($p = 0.11$). Specification (3) indicates that the stated importance of ESG for future investments outside the experiment is positively correlated with post-treatment *ESG Scores*. A one standard deviation increase in *ESG importance* is associated with a 13% increase in *ESG Score* compared to the average post-treatment score. These findings suggest a strong correlation between our survey-based risk measures and the bank's risk metrics, as well as between our revealed preference measure from the stock market game and both actual (planned) sustainability holdings and the future importance of ESG. This consistency adds further confidence in the reliability of our measures.

regular intervals for client-advisor contact, it can take some time until this information is available for all clients.

Table 42: Summary statistics of the main variables

	Mean	SD	Min	Max	N
ESG Score (prior)	21.53	5.62	15.99	34.30	74
ESG Score (post)	20.45	8.09	4.11	31.00	74
WTP	2,822.97	3,032.94	0.00	10,000.00	74
bank contact ESG	0.12	0.33	0.00	1.00	74
ESG importance	4.31	1.67	1.00	7.00	74
possess SRI	0.47	0.50	0.00	1.00	74
plan SRI	0.15	0.36	0.00	1.00	74
altruism	6.73	2.96	0.00	10.00	74
impact altruism	1.55	3.44	-5.00	10.00	74
warm glow altruism	5.19	3.50	0.00	10.00	74
lower SRI returns	3.30	1.88	0.00	7.00	74
lower SRI risk	3.74	2.09	0.00	7.00	74
ESG knowledge	0.74	0.44	0.00	1.00	74
risk preferences	6.23	2.10	2.00	10.00	74
risk preferences (financial instruments)	3.97	0.60	2.00	5.00	73
experience (financial products)	4.23	0.42	4.00	5.00	74
financial literacy	4.28	0.69	3.00	5.00	74
female	0.19	0.39	0.00	1.00	74
age	4.43	1.69	1.00	7.00	74
income	2.26	0.76	1.00	4.00	74
regular expenses	2.14	1.06	0.00	4.00	74
cash & stocks	3.20	0.91	1.00	4.00	74
other assets	2.58	1.16	1.00	5.00	74
wealth	2.58	0.70	1.00	3.00	74
ESG importance	4.31	1.67	1.00	7.00	74

Notes: *Table 42* presents summary statistics for the investor sample. We winsorize *ESG Score (prior)*, *ESG Score (post)*, and *financial literacy* on the 5% and the 95% level to reduce the impact of outliers. The bank had no *risk preferences (financial instruments)* data for one individual. For definitions of the variables, please consider Table 47.

Sample Characteristics: Our sample consists of experienced wealthy private investors. Each participant has interacted with at least one of the following investment types: bonds, funds, stocks, or private equity, with 23% additionally experienced in hedge funds, options, cum certificates, or knock-out certificates. Approximately 43% of the sample holds cash and stocks valued between 75,000 and 500,000 Euros, while another 43% holds assets exceeding 500,000 Euros, in contrast with the average German holding 88,600 Euros in cash and stocks (BVR, 2023). Furthermore, 47% of our participants have other assets like real estate valued between 100,000 Euros and 1,000,000 Euros, and 19% possess other assets valued over 1 million Euros, compared to the German average of 147,600 Euros (BIB, 2023; DESTATIS, 2023b). Taken together, more than 70% of the participants possess a net wealth of over 600,000 Euros. Besides being wealthy, these investors also show a high financial literacy, correctly answering 85% of our survey questions on average.[86] The demographic composition of our sample includes 19% females, with an average age of around 60 years. Median net incomes range from 2,000 to 5,000 Euros, compared to an

[86] Bucher-Koenen *et al.* (2017) show that in a representative survey of German households, 56% answered the first three of our financial literacy questions correctly, compared to 95% in our sample.

average income of 2,244 Euros in Germany in 2022 (DESTATIS, 2024), while average monthly regular expenses of our study subjects are between 1,000 and 2,000 Euros. Notably, 74% of participants were already familiar with ESG concepts before our study compared to 25% in the German population (IFNP, 2021). Last, participants generally perceive sustainable assets as slightly underperforming but displaying marginally lower risk compared to conventional assets. For a comprehensive overview of the data, please refer to Table 42.

5.4 Results

This paper has three main objectives: (i) to study correlations between altruistic preferences and sustainable investments; (ii) to measure the willingness to pay for sustainability information; (iii) to analyze how changes in return expectations affect revealed sustainability preferences. Unless stated otherwise, all analyses were pre-registered prior to data collection at the American Economic Association's registry for randomized controlled trials, AEARCTR-0009341.[87] It should be noted that within our framework, higher *ESG Scores* signify greater sustainability.

5.4.1 Altruistic Preferences and SRI

The literature establishes that investors value sustainable investments, as evidenced by Barreda-Tarrazona *et al.* (2011); Hartzmark and Sussman (2019); Pástor *et al.* (2020) among others. Our experimental findings with high-net-worth individuals align with these studies. Specifically, the average *ESG Score* from the baseline investment decision is 22.63, significantly exceeding the expected score of 18.33 for a randomly selected portfolio ($p < 0.01$), affirming findings from Braun *et al.* (2024b). In their study, financially inexperienced individuals in their 20s achieved a similar mean *ESG Score* of 22.68 in a comparable stock market game with exactly the same universe of stocks. Furthermore, the mean number of sustainable stocks held is 2.8, substantially higher than expected for a random selection ($p < 0.01$). Additionally, 62% of participants disclosed that they currently own or plan to acquire sustainable investments within the next three years. This raises a pertinent question: what drives sustainable investment behaviors among wealthy private investors?

Key studies identify altruistic preferences as pivotal for sustainable investment decisions (Riedl and Smeets, 2017; Bauer *et al.*, 2021; Gutsche *et al.*, 2023). Table 43 presents OLS regression results with *ESG Score* as the dependent variable for a sample of wealthy private investors, indicating a positive correlation between altruistic preferences and sustainable investments controlling for risk and return expectations of ESG investments, risk preferences, income,

[87] Despite pre-registering our analysis to include true sustainability portfolio holdings from our partnering bank, institutional constraints unfortunately prevented the bank from providing us with this data. We use individuals mean *ESG Scores* as the preferred specification as they are the richest regarding the degree of sustainability of an investor. In the appendix, you find all the results for alternative specifications of sustainability.

wealth, gender, age, and financial literacy.[88] Column (1) displays a statistically marginally non-significant positive correlation between *altruism* and *ESG Score* in the baseline decision of the stock market experiment ($p = 0.11$). The literature debates whether sustainable investments are primarily motivated by warm glow or impact altruism (Riedl and Smeets, 2017; Heeb *et al.*, 2023; Gutsche *et al.*, 2023). Understanding this distinction is crucial for developing investment strategies that align with investor motivations and for establishing regulatory frameworks that support genuine sustainability outcomes. If impact altruism predominates, providers of sustainable investment products should emphasize the tangible impacts of their products, such as contributions to sustainable development goals, emission reductions, or improved labor conditions. Consequently, there is no need for regulators to intervene as investors will identify truly sustainable investments, and greenwashing products will be discovered and removed from the market.[89] Conversely, if warm glow altruism is the main driver, issuers might focus on creating investment products that are labeled and perceived as sustainable, potentially providing more emotional satisfaction than actual impact (Gutsche *et al.*, 2023). Given the political efforts to use sustainable investments to finance the green transition of the economy (Commission, 2019a), regulators should prevent greenwashing by providing clear guidelines and labels that define what sustainable investments should entail. If ambiguity is removed, warm glow altruists can no longer rationalize questionable investments as sustainable and therefore might buy genuinely sustainable investments. Braun *et al.* (2024b) demonstrate that warm glow altruists are more likely to invest sustainably than non-altruists. However, in ambiguous information settings, they selectively acquire information telling them not to invest sustainably if they have an incentive to do so. As a consequence, they reduce their sustainable investments.

To differentiate these motives, we simultaneously include *warm glow altruism* and *impact altruism* in specification (2), which shows a positive correlation between *impact altruism* and *ESG Scores* ($p = 0.05$), with no significant correlation for *warm glow altruism* ($p = 0.32$). A one standard deviation rise in *impact altruism* increases the *ESG Score* by 7.5% compared to the average *ESG Score*. Controlling for *ESG knowledge* of the participants in specification (3), the correlation with *altruism* slightly decreases in significance ($p = 0.15$). Disentangling the effects of altruistic preferences, specification (4) confirms a significant correlation between *impact altruism* and *ESG Scores* ($p = 0.07$). Heterogeneity analysis reveals that the observed correlations are predominantly driven by investors knowledgeable about ESG.[90] Specification (5) confirms the correlation between *altruism* and *ESG Score* ($p = 0.03$).

[88] Table 50 (appendix) presents the outcomes for alternative specifications of the dependent variable. Although the results are not statistically significant, they are qualitatively consistent for the intensive margin, which assesses the degree of sustainability in investments. Conversely, we observe no effects on the extensive margin, which measures the probability of being a sustainable investor.

[89] Products might be removed due to reputational concerns of the issuer or because the demand for these products drops, leading to insufficient profits.

[90] Having *ESG knowledge* is strongly correlated with possessing or planning to possess sustainable investments (unreported tests), indicating that these individuals are predominantly experienced with ESG and/or possess a positive attitude towards ESG investments.

Table 43: Relation between sustainability and altruistic preferences

	(1) ESG Score	(2) ESG Score	(3) ESG Score	(4) ESG Score	(5) ESG Score	(6) ESG Score
altruism	1.078		0.981		1.654**	
	(0.656)		(0.664)		(0.718)	
lower SRI returns	-0.0187	0.0325	0.220	0.266	-0.643	-0.622
	(0.692)	(0.706)	(0.710)	(0.719)	(0.790)	(0.825)
lower SRI risk	0.286	0.0826	0.299	0.100	0.164	0.0968
	(0.751)	(0.804)	(0.746)	(0.801)	(0.815)	(0.872)
risk preferences	-0.948	-0.898	-1.091	-1.040	-1.574**	-1.532**
	(0.711)	(0.705)	(0.711)	(0.702)	(0.723)	(0.709)
income	0.945	0.878	0.836	0.774	0.107	0.0857
	(0.995)	(1.016)	(0.994)	(1.012)	(1.072)	(1.091)
wealth	-0.625	-0.554	-0.539	-0.471	0.140	0.142
	(0.755)	(0.757)	(0.751)	(0.751)	(0.931)	(0.942)
female	1.322	1.135	1.131	0.951	-0.695	-0.738
	(1.673)	(1.713)	(1.774)	(1.809)	(1.754)	(1.801)
financial literacy	0.998	0.822	0.910	0.740	1.361*	1.259
	(0.693)	(0.767)	(0.722)	(0.777)	(0.764)	(0.928)
age	-0.228	-0.0730	-0.216	-0.0652	-0.0814	-0.0158
	(0.624)	(0.636)	(0.599)	(0.606)	(0.791)	(0.815)
warm glow first	-2.139	-2.264	-1.954	-2.079	-2.532	-2.556
	(1.362)	(1.394)	(1.402)	(1.438)	(1.574)	(1.636)
impact altruism		1.610*		1.486*		2.046**
		(0.817)		(0.806)		(0.906)
warm glow altruism		0.924		0.815		1.860*
		(0.915)		(0.936)		(1.067)
ESG knowledge			2.296	2.262		
			(1.570)	(1.547)		
Constant	21.87***	21.96***	20.14***	20.26***	23.07***	23.12***
	(2.111)	(2.110)	(2.516)	(2.506)	(2.377)	(2.391)
N	74	74	74	74	55	55
adj. R squared	0.00766	0.00605	0.0256	0.0232	0.0760	0.0588
ESG knowledge	-	-	-	-	yes	yes

Notes: Table 43 demonstrates that altruistic preferences significantly correlate with *ESG Scores* across various OLS specifications. Specifications (1)-(4) analyze the full sample, while specifications (5) and (6) focus on individuals with pre-existing *ESG knowledge*, underscoring that the observed effects are primarily driven by those knowledgeable about ESG. The dependent variable in all specifications is the mean *ESG Score* from the baseline stock market game. Independent variables include three measures of altruistic preferences: (i) *altruism*, quantified by the donation amount in the dictator game; (ii) *warm glow altruism*, represented by donations in the "money-burning" dictator game; (iii) *impact altruism*, calculated as the difference between donations in the standard and "money-burning" dictator games. Control variables encompass *lower expected returns on SRI*, based on investors' responses to the statement *"I expect shares with a high ESG score to underperform conventional shares"*, scored from zero to seven; *lower perceived SRI risk*, responses to *"I expect shares with a high ESG score to have lower price fluctuations than conventional shares"*; *risk preferences*, indicating willingness to take risks on a scale of 0-10; *income*, personal net income per month; *wealth*, a categorical variable for the sum of an investors wealth; *female*, a binary indicator for gender; *financial literacy*, the count of correctly answered financial literacy quiz questions (0-5); *age*, the categorized age of an investor; and *ESG knowledge*, a binary indicator of whether an individual was familiar with ESG concepts prior to the experiment. The *Warm glow first* variable is binary, indicating whether the impure altruism task was conducted before the warm glow task. All non-binary variables are z-scored. Statistical significance levels are denoted by *, **, and *** for the 10, 5, and 1 percent levels, respectively.

Specification (6) displays a strong correlation of *impact altruism* with the sustainability indicator ($p = 0.03$), while also showing a statistically significant correlation of *warm glow altruism* with *ESG Score* ($p = 0.09$). These results emphasize that *impact altruism* is fundamental to sustainable investment preferences, and *warm glow altruism* plays a significant role predominantly among those familiar with ESG for wealthy private investors. The non significant results for *altruism* in specifications (1) and (3) are most likely due to the small sample size of 74 individuals. These findings are largely consistent with Braun *et al.* (2024b), who demonstrate a strong positive correlation between all three types of altruism and *ESG Scores* among young, inexperienced participants.

5.4.2 WTP for ESG Score information

After demonstrating that investors utilize ESG Scores, we next examine their willingness to pay for ESG Score information. In a hypothetical scenario, we observe a positive *WTP* for ESG Score information, averaging 2,822 points, which represents 5.6% of their available budget and is significantly greater than zero ($p < 0.01$). This finding aligns with Braun *et al.* (2024b), who, through an incentivized elicitation method, find that young participants are willing to allocate approximately 6.75% of their endowment to acquire sustainability information for trading in an incentivized stock market game. More broadly, this result corresponds with Baker *et al.* (2022), who document that index fund clients are prepared to pay, on average, 20 basis points higher fees for funds with an ESG mandate compared to identical funds without such a mandate, a noteworthy premium given the average expense ratio of these funds is 61 basis points.

Next, we examine what drives the WTP for ESG Score information. Guenster *et al.* (2022) show that the WTP for sustainable investments increase in an individual's level of altruism, as measured by a psychometric questionnaire. Heeb *et al.* (2023) argue that while investors demonstrate a significant willingness to pay for sustainable investments, their payments do not increase with greater impact, suggesting that *warm glow* rather than *impact altruism* motivates this willingness. Table 44 presents OLS regression results with *WTP* as the dependent variable, various measures of altruistic preferences as independent variables, and a set of controls. Specification (1) shows a positive, though not statistically significant, coefficient for *altruism* ($p = 0.32$). Specification (2) differentiates between various forms of altruistic preferences, revealing non-significant coefficients for both *impact altruism* ($p = 0.71$) and *warm glow altruism* ($p = 0.15$). Interestingly, there appears to be a negative correlation between *WTP* and *financial literacy*; across all specifications, a one standard deviation increase in *financial literacy* correlates with a reduction in *WTP* for ESG Risk information by 725 to 827 points, which is sizable given the mean *WTP* is around 2,800 points. This finding is exploratory but complements the results presented by Laudi *et al.* (2023), who illustrate that financial advisors often charge a premium for sustainable investments that is not justified by the effort, skill, or costs involved.

Table 44: Relation between the WTP, altruistic preferences, and financial literacy

	(1) WTP	(2) WTP
altruism	304.2	
	(342.8)	
lower SRI returns	147.2	114.0
	(400.8)	(407.7)
lower SRI risk	-534.4	-411.8
	(338.5)	(365.2)
risk preferences	468.5	443.5
	(359.3)	(345.6)
income	-132.8	-103.1
	(500.5)	(514.5)
wealth	870.0**	823.5*
	(434.2)	(453.3)
female	-350.3	-235.5
	(908.0)	(912.3)
financial literacy	-827.8***	-725.1**
	(308.6)	(329.1)
age	103.0	9.323
	(366.4)	(370.4)
warm glow first	1789.3***	1873.7***
	(660.1)	(675.5)
ESG knowledge	-900.0	-890.0
	(720.8)	(730.6)
impact altruism		154.0
		(428.0)
warm glow altruism		599.8
		(483.0)
Constant	1699.1*	1647.9*
	(902.5)	(901.2)
N	74	74
adj. R squared	0.145	0.147
controls	yes	yes

Notes: Table 44 presents evidence that altruistic preferences are not significantly related to the *WTP* for ESG Score information across various OLS specifications, while we find a significant negative correlation with *financial literacy.* The dependent variable across all specifications is the *WTP* for ESG Score information in a hypothetical scenario. We consider three measures of altruism as independent variables: (i) *altruism*, measured as the donation amount to charity in the dictator game; (ii) *warm glow altruism*, the donation amount in the "money-burning" dictator game; (iii) *impact altruism*, calculated as the difference between donations in the standard and "money-burning" dictator games. Controls include *lower expected returns on SRI*, the investor's response on a scale from zero to seven to the statement, *"I expect shares with a high ESG score to underperform conventional shares"*; *lower perceived SRI risk*, the response to *"I expect shares with a high ESG score to have lower price fluctuations than conventional shares"*; *risk preferences*, an indicator of willingness to take risks on a scale from 0 to 10; *income*, reported personal net income per month; *female*, a dummy variable for gender; *financial literacy*, the number of correctly answered questions in a financial literacy quiz (0-5); and *age*. *Warm glow first* is a dummy variable indicating if the impure altruism task preceded the warm glow task. *ESG knowledge* represents an indicator variable that is one if an individual knows about the concept of ESG before the experiment and zero otherwise. All non-indicator variables are z-scored. Significance levels are indicated by *, **, and *** at the 10, 5, and 1 percent levels, respectively.

Notably, this premium is more pronounced for sustainable investors who possess low financial literacy. Consequently, individuals who are financially less literate not only exhibit a higher willingness to pay but also face higher charges. Laudi *et al.* (2023) discuss that this form of price discrimination could be problematic by potentially harming consumer welfare and diminishing the long-term appeal of sustainable investments.

In summary, we observe a negative relationship between *WTP* and *financial literacy* and a non-significant relationship with altruistic preferences. This outcome diverges from the findings in Braun *et al.* (2024b), who observe a significant relationship with altruistic preferences and no relationship with *financial literacy* among young participants. Several factors may explain this discrepancy: First, *warm glow altruism* is at the margin of being significant at the 10% level, potentially due to a power issue given the small sample size of 74 investors. Second, due to time constraints, we use a hypothetical elicitation of the *WTP* in our sample, unlike in the student sample where the *WTP* had real consequences in a subsequent stock market game. Non-incentivized measures, although generally reliable, can sometimes overstate the true *WTP* and exhibit higher variance (List and Shogren, 1998). Lastly, the differing sample compositions—financially literate, experienced investors versus inexperienced young participants—might also contribute to the divergent results.

5.4.3 Return Expectations and Sustainable Investments

Standard economic theory posits that individuals respond to financial incentives (Markowitz, 1952); however, if investors are primarily motivated by social reasons to invest sustainably, financial incentives might not only fail to increase sustainable investments but could also negatively impact their intrinsic motivation to invest sustainably (Frey and Oberholzer-Gee, 1997; Mellström and Johannesson, 2008; Gneezy *et al.*, 2011). This may be particularly relevant given the demonstrated correlation between altruistic preferences and socially responsible investments. Døskeland and Pedersen (2016) apply the Levitt and List (2007) model, which assumes additively separable utility from financial and social benefits, on sustainable investments. Extending this model to include the potential negative externalities of positive financial incentives on the social utility derived from sustainable investments reveals four potential scenarios in response to risk-adjusted performance information about sustainable stocks.

When investors receive information indicating overperformance of sustainable stocks, they may (i) increase sustainable investments because the financial incentives align with their social goals, thus enjoying benefits from both financial and social aspects of their investments. Conversely, (ii) increased financial returns might undermine the intrinsic motivation for pro-social behavior, such as investing sustainably, potentially leading to unchanged or even reduced levels of sustainable investments as financial gains are offset by diminished social benefits. This aligns with Mellström and Johannesson (2008) findings on blood donation and Brodback *et al.* (2019) who argue that altruists' motivation to invest sustainably decreases when they expect higher returns from such investments. On the other hand, exposure to underperformance information about sustainable stocks (iii) leads to a reduction in sustainable investments if financial disadvantages significantly overshadow social benefits. Alternatively, (iv) such information might leave investment levels unchanged or even increase them if the perceived social benefits outweigh the financial drawbacks because the social value of such investments increases with the financial disutility.

To examine the causal impact of return expectations on sustainable investment behavior, we vary risk-adjusted return expectations by providing participants with exogenously assigned information about the overperformance or underperformance of sustainable investments. Table 51 in the appendix confirms successful randomization between treatments, showing no pre-treatment differences in *ESG Scores* or in variables correlated in Table 43 with sustainable investments such as altruistic preferences, *financial literacy*, and *risk preferences*. Minor pre-treatment discrepancies were observed in *female* and *ESG knowledge*. To address these and show the impact of within-subject reactions to overperformance and underperformance information, we apply an individual fixed effects regression model as the preferred specification:

$$ESG_{it} = \alpha + \beta_1 \cdot \text{overperformance}_{it} + \beta_2 \cdot \text{underperformance}_{it} + \gamma_i + \epsilon_{it} \tag{6}$$

where ESG_{it} is the mean *ESG Score* for individual i at time t, *overperformance* and *underperformance* are indicator variables for treatment assignment, γ_i represents individual fixed effects, and ϵ_{it} is the error term.

Table 45 presents the regression results. Column (1) shows that exposure to ESG overperformance (underperformance) information significantly increases (decreases) *ESG Scores* relative to the baseline portfolio allocation ($p < 0.01$), with changes of 22.8% (12.2%) respectively. This aligns with Braun *et al.* (2024b), who report increases (decreases) of 33.6% (14.2%) in a similar setting with a younger, less experienced demographic. Column (2) focuses on investors without prior *ESG knowledge*, showing a significant increase in *ESG Scores* following overperformance information ($p = 0.026$), but no significant change with underperformance information ($p = 0.36$).[91] This disparity is attributed to the lower baseline *ESG Scores* in this group.[92] These findings suggest that while overperformance information can enhance perceived risk-adjusted returns and thus encourage sustainable investments, underperformance information does not significantly deter investors. Specification (3) examines the effects on investors

[91]The heterogeneity analysis in specifications (2) and (3) is exploratory and illustrates how prior ESG knowledge influences individuals' responses to new information.

[92]T-tests confirm that baseline *ESG Scores* are significantly lower for investors without *ESG knowledge* compared to those with knowledge ($p < 0.01$) in combination with the small sample size. Correspondingly, pre-treatment return expectations for sustainable investments are significantly higher for individuals with *ESG knowledge* ($p = 0.01$), whereas there are no significant differences in pre-treatment risk expectations ($p = 0.60$).

with *ESG knowledge*, noting significant movements in *ESG Scores* for both overperformance ($p < 0.01$) and underperformance ($p = 0.07$) treatments.

Table 45: Relation between risk adjusted overperformance (underperformance) information and sustainable investing

	(1) ESG Score	(2) ESG Score	(3) ESG Score
overperformance	4.658***	6.555**	4.291***
	(1.286)	(2.707)	(1.448)
underperformance	-2.499**	-2.101	-2.714*
	(1.215)	(2.253)	(1.468)
Constant	20.45***	18.36***	21.17***
	(0.442)	(0.881)	(0.519)
N	148	38	110
fixed effects	yes	yes	yes
ESG knowledge	-	no	yes

Notes: Table 45 explores the impact of receiving risk-adjusted overperformance or underperformance information on sustainable investing using individual fixed effects regression. The dependent variable, *ESG Score*, is the mean ESG score of individual i at time $t \in \{1, 2\}$ (baseline and post-treatment portfolio allocations). *Overperformance* is an indicator variable set to one if individual i received the overperformance information at any time t, and zero otherwise. Similarly, *Underperformance* indicates whether individual i received underperformance information at any time t. Specifications (2) and (3) are exploratory; Specification (2) includes only investors with prior ESG knowledge, while Specification (3) is restricted to those familiar with the concept of ESG prior to the experiment. Significance levels are indicated by *, **, and *** for the 10, 5, and 1 percent levels, respectively.

Robustness analyses confirm that the effect of risk-adjusted performance information on investment behavior is largely consistent across various model specifications and dependent variables. Table 52 (appendix) provides empirical evidence that overperformance information exerts a persistent influence on investment choices. In contrast, the effects of underperformance information remain non-significant, which may be attributed to the limited statistical power of the study or the notion that a change in the extensive margin requires a stronger adjustment of the risk-adjusted return expectations. Table 53 in the appendix (not pre-registered) presents an OLS model with post-treatment *ESG Scores* (*ESG importance*) as the dependent variable, using an indicator for overperformance treatment participation. Specification (1) shows that investors receiving overperformance information hold portfolios with *ESG Scores* that are 42% higher than those receiving underperformance information ($p < 0.01$). This effect persists in Specification (2), which controls for the individual's *ESG Score* in the investment decision prior to the treatment. Specification (3) provides evidence that the treatment also influences future planned sustainable investments, showing a 0.44 standard deviation increase in the stated importance of ESG for future investment decisions ($p = 0.06$), an effect that is robust after controlling for baseline sustainability allocation.

These findings suggest that overall changes in the perceived performance of sustainable stocks can significantly alter both current and planned sustainable investment behaviors among wealthy private investors. Specifically, investors tend to increase their sustainable investments when financial incentives align with their social goals, thereby enjoying benefits from both the financial and social aspects of their investments. This finding is in line with Braun *et al.* (2024b). The evidence regarding the impact of underperformance information is somewhat ambiguous and warrants further investigation. Our preferred specification indicates a significant decrease in sustainable investments, consistent with the hypothesis that the financial disadvantages of investing sustainably may overshadow the associated social benefits. This effect is driven by individuals with ESG knowledge. Contrarily, our robustness analyses, employing various dependent variables, do not show significant reactions to underperformance information, likely due to the limited statistical power of this experiment. In contrast, Braun *et al.* (2024b) document a significant negative reaction to underperformance information across all specifications in their experiment involving over 600 young individuals. Nonetheless, we cannot dismiss the possibility that a change in the extensive margin to invest sustainably requires a more pronounced adjustment of risk-adjusted return expectations, aligning with the perspective that the social benefits of purchasing sustainable stocks might outweigh the financial disutility.

Table 46: The impact of the overperformance treatment (post-treatment ESG scores) on the demand for bank contact in relation to ESG

	(1) bank contact ESG	(2) bank contact ESG	(3) bank contact ESG
overperformance	0.0270		
	(0.0770)		
ESG Score (post)		0.0639*	0.0637*
		(0.0323)	(0.0343)
ESG Score (prior)			0.000951
			(0.0437)
Constant	0.108**	0.122***	0.122***
	(0.0518)	(0.0378)	(0.0381)
N	74	74	74
adj. R squared	-0.0122	0.0244	0.0107

Notes: Table 46 presents an OLS regression analysis showing a positive correlation between the desire to be contacted by the bank for further ESG information and post-treatment *ESG Scores*. We find no statistically significant relationship with the risk-adjusted return information treatment. The dependent variable, *bank contact ESG*, is an indicator set to one if the investor expresses a desire in the closing survey to receive more ESG-related information from the bank, and zero otherwise. *Overperformance* is an indicator variable, set to one for individuals in the overperformance treatment and zero for those in the underperformance treatment. *ESG Score (post)* and *ESG Score (prior)* represent the average ESG Scores of portfolio allocations after and before the treatment, respectively. Significance levels are indicated by *, **, and *** for the 10, 5, and 1 percent levels, respectively.

In the final stage of our analysis, we investigate whether performance information affects individuals' desire to be contacted by the bank for more ESG details. The rationale behind this is that if investors are more likely to invest sustainably in the stock market game, they might also be more likely to ask to be contacted by the bank to get more information on ESG and increase real-world sustainable investments. Table 46 presents the outcomes of an OLS regression, with *bank contact ESG* as the dependent variable and *overperformance* and post-treatment *ESG Scores* as key variables of interest. Column (1) indicates a positive but non-significant effect of *overperformance* on the inclination to be contacted by the bank ($p = 0.73$). Column (2), examining the direct impact of post-treatment *ESG Scores*, identifies a significant positive correlation ($p = 0.05$). Specifically, a one standard deviation increase in post-treatment *ESG Scores* elevates the probability of individuals wanting bank contact by 6.4 percentage points, noteworthy given the baseline demand for contact is 12 percentage points.[93] Column (3) reiterates these findings, adjusting for baseline *ESG Scores* to mitigate baseline sustainability preference disparities ($p = 0.07$). Overall, although the direct effect of overperformance information on requests for ESG-related contact is not significant, we observe a robust correlation between enhanced post-treatment ESG preferences and the increased desire for more information. This pattern suggests that changes in return expectations potentially influence not only sustainable investment behavior within the stock market game but also might extend to real-world sustainable investment decisions. Future research should further investigate this relationship in studies with greater statistical power, specifically focusing on measuring shifts in return expectations to clarify the underlying mechanisms.

5.5 Discussion and Conclusion

In this paper, we explore the influence of altruistic preferences and information on the performance of sustainable investments on the investment behavior of wealthy private investors. In collaboration with a German private bank, we conduct an online experiment involving an incentivized stock market game to elicit revealed preferences for socially responsible investments. Additionally, we measure various incentivized altruistic preferences and collect a comprehensive set of beliefs and demographic data, which we combine with administrative data from the bank's Investor Suitability Assessment. Our experimental design includes a controlled information treatment on the performance of sustainable stocks, allowing us to observe the investors' reactions within the stock market game. Furthermore, we assess investors' WTP for sustainability information, providing insights into the economic valuation of ESG data by affluent investors.

The aim of this paper is to extend common findings in the sustainable finance literature to a sample of wealthy private investors. Our findings align closely with those reported by Braun *et al.* (2024b) for a younger and financially inexperienced demographic, suggesting consistency

[93]Specifications (2) and (3) were not pre-registered. Despite this, they provide valuable insights, demonstrating a positive correlation between sustainable investment behavior observed in the stock market game and behaviors aligned with the intention to invest sustainably outside of the experimental setting. These findings highlight the consistency of investment preferences across different contexts, underscoring the relevance of the experimental results to real-world investor behavior.

across different investor profiles. Our main results are threefold. First, we show a marked preference for sustainable investments and establish a significant correlation between altruistic preferences and such investments. For investors lacking prior ESG knowledge, this correlation is primarily driven by impact altruism. Conversely, for those with ESG knowledge, both impact altruism and warm glow altruism influence their decisions. This extends existing literature that emphasizes the role of warm glow altruism (Gutsche and Ziegler, 2019; Heeb *et al.*, 2023; Gutsche *et al.*, 2023; Kleffel and Muck, 2023) by demonstrating the significance of impact altruism among wealthy private investors. This is encouraging for environmental sustainability, indicating that financial institutions should ensure their products genuinely generate impact. However, a significant portion of investors appear to be motivated by warm glow altruism, possibly preferring products that are labeled as sustainable but may lack substantive impact. This observation suggests that, to ensure sustainable investments truly contribute to environmental goals, regulators might need clear guidelines defining what constitutes impact investments. The Sustainable Finance Disclosure Directive (Commission, 2019b) aims to aid European investors in making informed decisions by setting disclosure requirements for ESG metrics for firms and financial products. Nevertheless, it remains challenging for investors to identify genuine impact investments.[94] Braun *et al.* (2024a) highlight difficulties investors face under current regulations in identifying true impact investments, as there is a tendency to rationalize metrics that align with financial gains as impactful. Future research should delve deeper into the effects of different forms of altruism on impact investments and explore the effectiveness of marketing campaigns that emphasize the emotional versus real-world impact of sustainable investments, assessing which resonates more with various types of investors.

Second, our findings reveal that investors, on average, allocate 5.6% of their budget to acquire ESG Score information for their potential investments, which is in line with Braun *et al.* (2024b). Additionally, we observe that financially illiterate investors are willing to pay significantly more than financially literate investors. Laudi *et al.* (2023) demonstrate that financial advisors often charge higher fees to financially illiterate individuals seeking sustainable investments. Such pricing strategies could potentially reduce consumer welfare and diminish the long-term attractiveness of sustainable investments for these groups. Consequently, policymakers might consider monitoring fees for information sources that assist investors in making sustainable investment decisions to prevent disadvantageous price discrimination.

Third, we observe that participants' investment behavior is highly responsive to information about the risk-adjusted performance of sustainable investments, demonstrating a sensitivity to financial motives. Exposure to information suggesting overperformance (underperformance) of sustainable stocks results in an increase (decrease) in ESG Scores relative to the same investors' baseline portfolio allocations of 22.8% (12.2%). These findings align with Døskeland and Ped-

[94] See Kölbel *et al.* (2020) for an overview on what impact investments entail.

ersen (2016); Giglio *et al.* (2023); Braun *et al.* (2024b), which indicate that risk and return expectations are crucial determinants of sustainable investments. Therefore, implementing tax discounts on returns from sustainable investments or subsidies on the costs associated with these investments could be effective strategies for policymakers to foster sustainable investment activity. Future research might explore whether such responsiveness also applies to impact investments. Moreover, further studies should delve deeper into the effects of negative incentives, enhancing our understanding of how financial outcomes influence sustainable investments.

More broadly, this study utilizes revealed sustainability preferences from a stock market game. Despite implementing several measures to minimize experimenter demand effects and hypothetical bias (List and Shogren, 1998; List, 2001; Harrison, 2006; Harrison and Rutström, 2008; Haaland *et al.*, 2023), it cannot be entirely ruled out that investors might behave differently when they do not feel observed (Roethlisberger and Dickson, 2003), with higher stakes (Kirchler *et al.*, 2018), in their usual broker setting, or during consultations with their financial advisors. Future research could benefit from a richer sample size, allowing for a more detailed exploration of these dynamics. Taken together, this study shows several factors influencing wealthy investors' sustainable investment decisions. By matching experimental data from an obfuscated study with brokerage data or sustainability preferences from the Investor Suitability Assessment, subsequent studies could yield further insights into investor behaviors.

5.6 Appendix

Table 47: Variable definitions

Variable	Measure
ESG Score	The average ESG Score of an investor's baseline and post-treatment portfolio choice.
# sustainable	Denotes the number of sustainable stocks of an investor in the baseline and post-treatment portfolio choice.
# sustainable d.	Dummy variable equal to one if the participant invests in more than two sustainable stocks in a given portfolio allocation.
WTP	The amount of points an investor is willing to spend to obtain the ESG scores in a hypothetical scenario.
bank contact ESG	Indicator set to one if the investor expresses a desire in the closing survey to receive more ESG-related information from the bank, and zero otherwise.
ESG importance	Measures investor's agreement with the statement *"The sustainability of my future investments is very important to me."* on a seven-point Likert scale.
overperformance	Dummy for participation in the overperformance treatment, 1 if received, 0 otherwise.
underperformance	Dummy for participation in the underperformance treatment, 1 if received, 0 otherwise.
altruism	The amount of the donation to the charity in the standard dictator game (between 0 and 10 Euros).
impact altruism	The difference in the amount of the donation to the charity in the standard dictator game and the "money-burning" dictator game.
warm glow altruism	The amount of the donation to the charity in the "money-burning" dictator game (between 0 and 10 Euros).
warm glow first	Dummy variable equal to one if the investor participated in the impure altruism dictator game task before the warm glow dictator game task.
lower SRI returns	The investor´s response to the statement *"I expect shares with a high ESG score to underperform conventional shares."* on a scale from 0 (does not apply at all) to 7 (applies completely).
lower SRI risk	The investor´s response to the statement *"I expect shares with a high ESG score to have lower price fluctuations than conventional shares."* on a scale from 0 to 7.
ESG knowledge	Dummy variable equal to one if the investor knows the concept of ESG before the study.
possess SRI	Indicator, 1 if owning sustainable investments prior to the experiment, 0 otherwise.
plan SRI	Indicator, 1 if planning to make sustainable investments within the next three years, 0 otherwise.
possess or plan SRI	Indicator, 1 if currently owns or plans to invest in sustainable investments within three years, 0 otherwise.

Table 47 (continued):

Variable	Measure
risk preferences	Indicates investors' risk attitude as measured in the question *"Please tell me, in general, how willing or unwilling you are to take risks, using a scale from 0 to 10,"*.
risk preferences (financial instruments)	Indicates investors' answer to the question *"What level of risk are you prepared to take as part of your investment strategy for individual securities investments?"*, rated on a scale from one to five.
experience (financial products)	Indicates investors' experience and knowledge in various categories of securities, rated from one to five.
riskiest asset class	Indicator variable set to one for investors experienced in the riskiest asset classes (e.g., options, cum certificates, hedge funds).
financial literacy	Number of correctly answered questions from the financial literacy quiz (0-5).
female	Dummy variable, 1 if the investor is female, 0 otherwise.
age	Categorical variable representing investor's age group: (1) 21-30, (2) 31-40, (3) 41-50, (4) 51-60, (5) 61-70, (6) 71-80, (7) 81+ years.
income	Monthly net income categories: (1) < 2,000 Euros, (2) 2,001 - 5,000 Euros, (3) > 5,001 Euros.
regular expenses	Monthly expenses categories: (0) None, (1) < 1,000 Euros, (2) 1,001 - 2,000 Euros, (3) 2,001 - 5,000 Euros, (4) > 5,000 Euros.
cash & stocks	Cash and stock holdings categories: (1) < 25,000 Euros, (2) 25,001 - 75,000 Euros, (3) 75,001 - 500,000 Euros, (4) > 500,000 Euros.
other assets	Other assets like real estate: (1) None, (2) < 100,000 Euros, (3) 100,001 - 1,000,000 Euros, (4) 1,000,001 - 5,000,000 Euros, (5) > 5,000,000 Euros.
wealth	Total wealth categories: (1) < 175,000 Euros, (2) 175,001 - 600,000 Euros, (3) > 600,000 Euros.

Notes: Table 47 defines all relevant variables of the wealthy private investor sample.

Table 48: Relation between risk preferences, willingness to take losses with financial products, and experience with financial products

	(1) risk preferences	(2) risk preferences	(3) risk preferences
risk preferences (financial instruments)	0.391***		0.335***
	(0.110)		(0.111)
experience riskiest asset class		0.732***	0.546**
		(0.265)	(0.267)
Constant	-0.00503	-3.095***	-2.308**
	(0.109)	(1.125)	(1.130)
N	73	74	73
adj. R squared	0.139	0.0835	0.176

Notes: Table 48 demonstrates a significant correlation between the bank's risk attitude (experience) measures and the global preference survey risk measure (Falk *et al.*, 2018). The dependent variable, *risk preferences*, is derived from investors' responses to the question, *"Please tell me, in general, how willing or unwilling you are to take risks, using a scale from 0 to 10,"* where higher values indicate greater willingness to take risks (Falk *et al.*, 2018). The variable of interest, *risk preferences (financial instruments)*, reflects investors' stated risk tolerance for individual securities investments, rated on a scale from one to five. *riskiest asset class* is an indicator variable valued at one for investors experienced with the highest-risk asset classes (options, cum certificates, knock-out certificates, or hedge funds) and zero otherwise. Both, *risk preferences* and *risk preferences (financial instruments)* are z-scored. Whenever indicated, *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively.

Table 49: Relation between reporting to (plan to) possess sustainable investments and revealed sustainability preferences

	(1) ESG Score (prior)	(2) ESG Score (prior)	(3) ESG Score (post)
possess SRI	1.670		
	(1.421)		
plan SRI	3.799*		
	(1.992)		
possess or plan SRI		2.179	
		(1.351)	
ESG importance			2.701***
			(0.939)
Constant	20.18***	20.18***	20.45***
	(1.094)	(1.087)	(0.893)
N	74	74	74
adj. R squared	0.0255	0.0224	0.0991

Notes: Table 49 presents evidence of a correlation between revealed sustainability preferences in the stock market game and both existing and planned sustainable investments in real portfolio holdings, as well as the stated importance of ESG for future investments. Specifications (1) and (2) use the dependent variable *ESG Score (prior)*, which is the mean ESG Score of the five stocks selected during the baseline portfolio allocation. The dependent variable in Specification (3), *ESG Score (post)*, represents the mean *ESG Score* from the post-treatment portfolio allocation. The variables of interest include: *possess SRI*, an indicator variable that is one if an investor reports owning sustainable investments prior to the experiment and zero otherwise; *plan SRI*, which is one if an investor does not currently possess sustainable investments but plans to make such investments within the next three years; *possess or plan SRI*, an indicator variable that is one if an investor either currently owns or plans to invest in sustainable investments within the next three years. *ESG importance* measures agreement with the statement *"The sustainability of my future investments is very important to me."* on a seven-point Likert scale. Significance levels *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 50: Relation between beeing a sustainable investor and altruistic preferences

	(1) # sustainable	(2) # sustainable	(3) # sustainable d.	(4) # sustainable d.
altruism	0.165		-0.0350	
	(0.154)		(0.0680)	
lower SRI returns	0.0435	0.0545	-0.00770	0.000332
	(0.164)	(0.165)	(0.0683)	(0.0694)
lower SRI risk	0.0594	0.0136	0.0442	0.0139
	(0.158)	(0.163)	(0.0605)	(0.0627)
risk preferences	-0.240	-0.229	-0.0517	-0.0452
	(0.172)	(0.169)	(0.0626)	(0.0630)
income	0.147	0.133	0.118	0.110
	(0.268)	(0.271)	(0.109)	(0.109)
wealth	-0.0312	-0.0152	-0.0453	-0.0340
	(0.185)	(0.187)	(0.0799)	(0.0795)
female	0.231	0.189	0.172	0.143
	(0.412)	(0.423)	(0.158)	(0.160)
financial literacy	0.232	0.193	0.0458	0.0203
	(0.174)	(0.181)	(0.0650)	(0.0688)
age	-0.0808	-0.0463	-0.0155	0.00768
	(0.144)	(0.152)	(0.0642)	(0.0687)
warm glow first	-0.351	-0.380	-0.0899	-0.110
	(0.313)	(0.315)	(0.123)	(0.123)
ESG knowledge	0.662*	0.655*	0.261*	0.258*
	(0.345)	(0.341)	(0.137)	(0.136)
impact altruism		0.273		0.00943
		(0.190)		(0.0828)
warm glow altruism		0.116		-0.0996
		(0.212)		(0.0897)
Constant	2.247***	2.274***	0.242	0.255
	(0.599)	(0.596)	(0.199)	(0.196)
N	74	74	74	74
adj. R squared	0.0199	0.0178	-0.0277	-0.00820

Notes: Table 50 illustrates that altruistic preferences are not significantly associated with *# sustainable* or *# sustainable d.* across various OLS specifications. The dependent variable in specifications (1)-(2), *# sustainable*, represents the count of sustainable stocks chosen during the baseline stock market game. In specifications (3)-(4), the dependent variable is *# sustainable d.*, a binary indicator set to one if an investor selects more than two sustainable stocks in the baseline portfolio allocation, and zero otherwise. Independent variables include three measures of altruistic preferences: (i) *altruism*, quantified by the donation amount in the dictator game; (ii) *warm glow altruism*, represented by donations in the "money-burning" dictator game; (iii) *impact altruism*, calculated as the difference between donations in the standard and "money-burning" dictator games. Control variables encompass *lower expected returns on SRI*, based on investors' responses to the statement *"I expect shares with a high ESG score to underperform conventional shares"*, scored from zero to seven; *lower perceived SRI risk*, responses to *"I expect shares with a high ESG score to have lower price fluctuations than conventional shares"*; *risk preferences*, indicating willingness to take risks on a scale of 0-10; *income*, personal net income per month; *female*, a binary indicator for gender; *financial literacy*, the count of correctly answered financial literacy quiz questions (0-5); *age*, the categorized age of an investor; and *ESG knowledge*, a binary indicator of whether an individual was familiar with ESG concepts prior to the experiment. The *Warm glow first* variable is binary, indicating whether the impure altruism task was conducted before the warm glow task. All non-binary variables are z-scored. Statistical significance levels are denoted by *, **, and *** for the 10, 5, and 1 percent levels, respectively.

Table 51: No significant differences in personal characteristics between overperformance and underperformance treatment

	Overp-performance	Under-performance	Diff.	Std. Error	Obs.
ESG Score (prior)	21.48	21.58	0.10	1.32	74
WTP	2548.65	3097.30	548.65	707.07	74
altruism	6.92	6.54	-0.38	0.69	74
impact altruism	1.89	1.22	-0.68	0.80	74
warm glow altruism	5.03	5.35	0.32	0.82	74
lower SRI returns	3.14	3.46	0.32	0.44	74
lower SRI risk	3.76	3.73	-0.03	0.49	74
ESG knowledge	0.84	0.65	-0.19*	0.10	74
risk preferences	6.57	5.89	-0.68	0.48	74
risk preferences (financial instruments)	4.06	3.89	-0.16	0.14	73
experience (financial products)	4.22	4.24	0.03	0.10	74
financial literacy	4.35	4.22	-0.14	0.16	74
female	0.27	0.11	-0.16*	0.09	74
age	4.54	4.32	-0.22	0.39	74
income	2.38	2.14	-0.24	0.18	74
regular expenses	2.32	1.95	-0.38	0.25	74
cash & stocks	3.27	3.14	-0.14	0.21	74
other assets	2.59	2.57	-0.03	0.27	74
wealth	2.62	2.54	-0.08	0.16	74

Notes: *Table 51* presents summary statistics for the investor sample separately for investors receiving the overperformance treatment and investors receiving the underperformance treatment. Most importantly, the results show no differences in *ESG Scores* in the baseline decision. Furthermore, the distribution of all other variables is what would be expected. If anything, we see marginal significant differences for *female* and *ESG knowledge*.

Table 52: Relation between risk adjusted overperformance (underperformance) information and sustainable investing

	(1) # sustainable	(2) # sustainable d.
overperformance	1.187***	0.329***
	(0.213)	(0.0780)
underperformance	0.00244	0.0498
	(0.219)	(0.0889)
Constant	2.770***	0.527***
	(0.150)	(0.0588)
N	148	148
fixed effects	yes	yes

Notes: Table 45 explores the impact of receiving risk-adjusted overperformance or underperformance information on sustainable investing using individual fixed effects regression. The dependent variable in Specification (1), *# sustainable*, denotes the number of sustainable stocks of individual i at time $t \in \{1, 2\}$ (baseline and post-treatment portfolio allocations). The dependent variable in Specification (2), *# sustainable d.*, is one if investor i at time $t \in \{1, 2\}$ (baseline and post-treatment portfolio allocations) selects more than two sustainable stocks. *Overperformance* is an indicator variable set to one if individual i received the overperformance information at any time t, and zero otherwise. Similarly, *Underperformance* indicates whether individual i received underperformance information at any time t. Significance levels are indicated by *, **, and *** for the 10, 5, and 1 percent levels, respectively.

Table 53: Effect of performance information on (future) sustainable investments

	(1) ESG Score (post)	(2) ESG Score (post)	(3) ESG importance	(4) ESG importance
overperformance	7.058***	7.100***	0.437*	0.441*
	(1.701)	(1.614)	(0.228)	(0.223)
ESG Score (prior)		0.432***		0.0410**
		(0.147)		(0.0203)
Constant	16.92***	7.610**	-0.218	-1.102**
	(1.270)	(3.329)	(0.162)	(0.447)
N	74	74	74	74
adj. R squared	0.182	0.263	0.0351	0.0761

Notes: Table 53 illustrates the significant impact of performance information on investors' (future) sustainable investment decisions. The dependent variable, *ESG Score (post)*, represents the mean ESG Score of investors' portfolios after receiving treatment information. Additionally, *ESG importance* measures investors' agreement with the statement *"The sustainability of my future investments is very important to me"*. The primary variable of interest, *overperformance*, is an indicator variable set to one if an investor receives the overperformance treatment, and zero if the underperformance treatment is received. Both *ESG importance* and *ESG Score (prior)* are z-scored. Significance levels are denoted by *, **, and *** at the 10, 5, and 1 percent levels, respectively.

Firm 100

Sector 2

1. Firm Information and Key Figures

January 2019	
Dividend[1] [Euro]	3,07
Price Earnings ratio[2]	20,68
Volatility[3] (1. J) [%]	23,00
Share Price [Euro][4]	53,68

January 2020	
Equity *Value*[5] [Euro]	991.457.040
# Employees[6]	36.000
Revenue[7] [Euro]	1.473.060.960
Debt Equity Ratio[8] [%]	403

2. Past Performance (January 2019)

Period	1 Month	3 Months	1 Year	3 Years
Past Performance[9]	-6,96%	-9,08%	-1,92%	-33,87%
Past Performance rel. to MSCI World[10]	0,50%	4,49%	6,64%	-56,26%

3. ESG Risk Informationen (2020)

ESG Risk Score[11]

17,8 Low Risk

Very Low	Low	Medium	High	Very High
0-10	10-20	20-30	30-40	40+

Components of ESG Risk *Scores*[12]

	Very Low 0 - 3,3	Low 3,3 - 6,6	Medium 6,6 - 10	High 10 - 13,3	Very High 13,3+	
Environment Risk Score						0,1
Social Risk Score						10,1
Governance Risk Score						7,6

Figure 23: Fact sheet of stock 100

Notes: This figure illustrates a translated version of a typical fact sheet from the stock market game used in the experiment. Participants could view this fact sheet by clicking on a stock's name within the trading interface. The fact sheet includes data from 2019, such as dividend yield, price-to-earnings (PE) ratio, volatility, share price, and past performance. Additionally, it features 2020 information like equity value, number of employees, turnover, and debt ratio. It also provides details about the ESG Risk Score, enhancing participants' ability to make informed investment decisions based on both financial metrics and sustainability factors.

Chapter 6

6 Conclusion

This dissertation builds on the rapidly growing literature on sustainable finance and links it to insights from experimental and behavioral economics. In four essays, my co-authors and I explore different aspects of the overarching question of what drives sustainable investment decisions. We analyze investors' reactions to (ambiguous) information, examine the consequences of (motivated) beliefs on investment decisions as well as perceptions of an appropriate definition of sustainable investments, and study the relationship between preferences, demographics, and sustainable investments. To this end, we conduct three incentivized online experiments with various demographic groups.

The first essay analyzes the role of preferences, incentives, and information for sustainable investment decisions. In doing so, the essay contributes to the overall question by identifying various factors that promote or hinder sustainable investments. The essay has two main findings: First, moral and financial considerations play a role in sustainable investment decisions. Investors respond to information about the moral implications of sustainable investments as well as to (perceived) financial incentives. In addition, altruistic individuals are more likely to invest sustainably. Second, the results show that investors opportunistically acquire information about the moral implications of sustainable investments. People choose to read the information that is most consistent with their financial incentives. Since investors subsequently use this information to make investment decisions, biased information acquisition prevents the efficient transmission of existing sustainability preferences.

The second essay relates to the overall question by showing what shapes investors' perceptions of what an appropriate definition of sustainable investments should entail. The paper offers three different definitions of sustainability and provides pro and con arguments for each. The results present causal evidence that investors' perceptions of the most appropriate sustainability definition are influenced by motivated beliefs. These motivated beliefs are shaped by investors' return expectations as well as their knowledge about the sustainability characteristics of their portfolio holdings within the experiment. By using open-ended responses, the study shows that participants use and interpret the same arguments differently to justify their prior decisions.

The third essay contributes to the overall question by using Machine Learning techniques to identify the defining characteristics of the next generation of retail investors and to examine the predictive power of behavioral, demographic, and portfolio characteristics in predicting sustainable investor types. Using K-Means clustering, data-driven personas for sustainable and non-sustainable investors are developed. Subsequent analysis employing Random Forest, XG-Boost, K-Nearest Neighbor, and Logistic Regression reveals that while personal characteristics such as financial literacy, education, and income modestly predict sustainable investing, the inclusion of portfolio characteristics significantly improves model performance.

The final essay connects to the question of what drives sustainable investing by analyzing the characteristics of wealthy individuals with a propensity for sustainable investing. Investors with a net worth between 100,000 USD and 1,000,000 USD are of particular interest as they account for almost 40% of private financial assets (Credit Suisse 2023). However, most existing studies focus on demographic groups with lower wealth levels. The results show that wealthy individuals are willing to pay for information that helps them make sustainable investment decisions. Moreover, altruistic preferences are correlated with sustainable investment decisions. Finally, the findings demonstrate that wealthy private investors are responsive to information about the risk-adjusted performance of sustainable investments.

In summary, the results shed light on various factors that influence sustainable investment decisions. As shown in Chapters 2 and 3, motivated beliefs can prevent investors from investing sustainably. Since motivated beliefs operate in an environment of ambiguous information, clear rules and guidelines are required. To date, there is no universal definition of sustainable investments, and the broad European approach is to complicated for investors, as shown in Chapter 3. Moreover, Chapters 2 and 5 provide evidence that while investors care about the moral implications of their investments, a significant proportion of these investors are warm glow altruists who are susceptible to greenwashing. Therefore, transparent data, consistent ESG ratings, and clear information about the moral consequences of SRI are key to enabling optimal capital allocation and preventing strategic self-deception. Future regulations should simplify the definition of sustainability for investment purposes to enable the effective transmission of sustainability preferences. Policymakers seeking to attract private capital to finance the green transition can use the findings to improve regulations, researchers can incorporate motivated beliefs into their economic models, and financial institutions may highlight the moral implications of sustainable investments to attract investors who want to finance the green transition. Furthermore, this dissertation shows that financial considerations play a role in sustainable investment decisions. Chapters 2, 3 and 5 illustrate that investors respond significantly to (perceived) financial incentives. Policymakers could therefore use tax rebates on returns or subsidies on the costs of sustainable investments to attract more private capital to finance the greening of the economy. In addition, researchers can be assured that the basic economic principles of Markowitz (1952) also apply in the context of sustainable investments. Finally, researchers, regulators, and financial institutions can use the results of this dissertation to understand who the sustainable investors are. Chapters 1, 3, and 5 identify factors that influence sustainable investing for different demographic groups. In addition, Chapter 4 constructs data-driven investor personas and uses Machine Learning models to predict sustainable investor types. These findings can be used by financial institutions to develop new products that help finance the green transition. Moreover, regulators and researchers can use the findings to integrate sustainable investors into their regulatory frameworks and economic models, thereby improving their understanding of how investors interact in financial markets.

References

Aas, K., Jullum, M. and Løland, A. (2021). Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. *Artificial Intelligence*, **298**, 1–24.

Abdi, H. and Williams, L. J. (2010). Principal component analysis. *WIREs Computational Statistics*, **2** (4), 433–459.

Akre, V., Rajan, A., Ahamed, J., Al Amri, A. and Al Daisi, S. (2019). Smart Digital Marketing of Financial Services to Millennial Generation using emerging technological tools and buyer persona. In *2019 Sixth HCT Information Technology Trends (ITT)*, pp. 120–125.

Ambuehl, S. (2024). An experimental test of whether financial incentives constitute undue inducement in decision-making. *Nature Human Behaviour*, **8** (5), 835–845.

Amelio, A. and Zimmermann, F. (2023). Motivated Memory in Economics—A Review. *Games*, **14** (1), 1–15.

An, J., Kwak, H., Jung, S.-g., Salminen, J. and Jansen, B. J. (2018). Customer segmentation using online platforms: isolating behavioral and demographic segments for persona creation via aggregated user data. *Social Network Analysis and Mining*, **8** (1), 54.

Anderson, A. and Robinson, D. T. (2022). Financial Literacy in the Age of Green Investment*. *Review of Finance*, **26** (6), 1551–1584.

Andreoni, J. (1989). Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence. *Journal of Political Economy*, **97** (6), 1447–1458.

— (1990). Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *The Economic Journal*, **100** (401), 464.

Arian, A., Meiyu Pan, M. and Chiu, Y.-C. (2021). Personas: A Market Segmentation Approach for Transportation Behavior Change. *Transportation Research Record*, **2675** (11), 172–185.

Ashabi, A., Sahibuddin, S. B. and Salkhordeh Haghighi, M. (2020). The Systematic Review of K-Means Clustering Algorithm. In *2020 The 9th International Conference on Networks, Communication and Computing*, Tokyo Japan: ACM, pp. 13–18.

Athey, S., Tibshirani, J. and Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, **47** (2), 1148–1178.

Auzepy, A., Bannier, C. E. and Gärtner, F. (2024). Looking beyond ESG preferences: The role of sustainable finance literacy in sustainable investing.

Avramov, D., Cheng, S., Lioui, A. and Tarelli, A. (2022). Sustainable investing with ESG rating uncertainty. *Journal of Financial Economics*, **145** (2), 642–664.

Awad, N., Shala, G., Deng, D., Mallik, N., Feurer, M., Eggensperger, K., Biedenkapp, A., Vermetten, D., Wang, H., Doerr, C., Lindauer, M. and Hutter, F. (2020). Squirrel: A Switching Hyperparameter Optimizer.

Bahmani, B., Moseley, B., Vattani, A., Kumar, R. and Vassilvitskii, S. (2012). Scalable K-Means++.

Baker, M., Egan, M. L. and Sarkar, S. K. (2022). How Do Investors Value ESG?

Barreda-Tarrazona, I., Matallín-Sáez, J. C. and Balaguer-Franch, M. R. (2011). Measuring Investors' Socially Responsible Preferences in Mutual Funds. *Journal of Business Ethics*, **103** (2), 305–330.

Bauer, K., von Zahn, M. and Hinz, O. (2023). Expl(AI)ned: The Impact of Explainable Artificial Intelligence on Users' Information Processing. *Information Systems Research*, **34** (4), 1582–1602.

Bauer, R., Ruof, T. and Smeets, P. (2021). Get Real! Individuals Prefer More Sustainable Investments. *Review of Financial Studies*, **34** (8), 3976–4043.

— and Smeets, P. (2015). Social identification and investment decisions. *Journal of Economic Behavior & Organization*, **117**, 121–134.

Becker, G. M., DeGroot, M. H. and Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral science*, **9** (3), 226–232.

Begenau, J. and Siriwardane, E. N. (2024). Fee Variation in Private Equity. *The Journal of Finance*, **79** (2), 1199–1247.

Bekkerman, R., Cohen, M. C., Kung, E., Maiden, J. and Proserpio, D. (2023). The Effect of Short-Term Rentals on Residential Investment. *Marketing Science*, **42** (4), 819–834.

Berg, F., Kölbel, J. F. and Rigobon, R. (2022). Aggregate Confusion: The Divergence of ESG Ratings. *Review of Finance*, **26** (6), 1315–1344.

Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *The Journal of Machine Learning Research*, **13** (0), 281–305.

Betermier, S., Calvet, L. E. and Sodini, P. (2017). Who Are the Value and Growth Investors? *The Journal of Finance*, **72** (1), 5–46.

BIB (2023). Zahl der Privathaushalte und durchschnittliche Haushaltsgröße in Deutschland (1871-2022).

Billio, M., Costola, M., Hristova, I., Latino, C. and Pelizzon, L. (2021). Inside the ESG ratings: (Dis)agreement and performance. *Corporate Social Responsibility and Environmental Management*, **28** (5), 1426–1445.

—, Fitzpatrick, A. C., Latino, C. and Pelizzon, L. (2024). Unpacking the ESG Ratings: Does One Size Fit All?

Board, S. F. A. (2024). Abschlussempfehlung zur Einführung einer ESG-Skala für Finanzprodukte vom Februar 2024.

Bogan, V. (2008). Stock Market Participation and the Internet. *Journal of Financial and Quantitative Analysis*, **43** (1), 191–211.

Bogousslavsky, V., Fos, V. and Muravyev, D. (2024). Informed Trading Intensity. *The Journal of Finance*, **79** (2), 903–948.

Bolton, P. and Kacperczyk, M. (2021). Do investors care about carbon risk? *Journal of Financial Economics*, **142** (2), 517–549.

— and — (2023). Global Pricing of Carbon-Transition Risk. *The Journal of Finance*, **78** (6), 3677–3754.

— and — (2024). Are Carbon Emissions Associated with Stock Returns? Comment. *Review of Finance*, **28** (1), 107–109.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, **30** (7), 1145–1159.

Braun, F., Gill, A. and Hett, F. (2024a). Shifting Grounds: The Influence of Motivated Beliefs on Investor Convictions about Sustainability.

—, — and — (2024b). What Drives Socially Responsible Investments: The Role of Preferences, Incentives, and Information.

—, — and — (2024c). What Motivates Wealthy Private Investors to Hold Sustainable Investments?

Breiman, L. (1996). Bagging predictors. *Machine Learning*, **24** (2), 123–140.

— (2001). Random Forests. *Machine Learning*, **45** (1), 5–32.

—, Friedman, J., Stone, C. J. and Olshen, R. A. (eds.) (1984). *Classification and regression trees*. Boca Raton, Fla.: Chapman & Hall/CRC, 1st edn.

Briere, M. and Ramelli, S. (2021). Green Sentiment, Stock Returns, and Corporate Behavior.

Brodback, D., Guenster, N. and Mezger, D. (2019). Altruism and egoism in investment decisions. *Review of Financial Economics*, **37** (1), 118–148.

Brusco, M. J. and Cradit, J. D. (2001). A variable-selection heuristic for K-means clustering. *Psychometrika*, **66** (2), 249–270.

Bucher-Koenen, T., Lusardi, A., Alessie, R. and van Rooij, M. (2017). How Financially Literate Are Women? An Overview and New Insights. *Journal of Consumer Affairs*, **51** (2), 255–283.

BVR (2023). 31. Mai 2023: Die deutschen Sparer trotzen Inflation und Energiekrise - vorerst?

Bénabou, R. and Tirole, J. (2002). Self-Confidence and Personal Motivation*. *The Quarterly Journal of Economics*, **117** (3), 871–915.

— and — (2016). Mindful Economics: The Production, Consumption, and Value of Beliefs. *Journal of Economic Perspectives*, **30** (3), 141–164.

Calders, T. and Jaroszewicz, S. (2007). Efficient AUC Optimization for Classification. In J. N. Kok, J. Koronacki, R. Lopez De Mantaras, S. Matwin, D. Mladenič and A. Skowron (eds.), *Knowledge Discovery in Databases: PKDD 2007*, vol. 4702, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 42–53.

Cartry, C., Lévêque, C., Jacamon, V., Trüby, J., Lorentz, B., Philip, P. and Shirizadeh, B. (2023). Financing the green energy transition: A US$50-trillion catch.

Chapman, J., Dean, M., Ortoleva, P., Snowberg, E. and Camerer, C. (2023). Econographics. *Journal of Political Economy Microeconomics*, **1** (1), 115–161.

Charness, G., Gneezy, U. and Halladay, B. (2016). Experimental methods: Pay one or pay all. *Journal of Economic Behavior & Organization*, **131**, 141–150.

Chava, S. (2014). Environmental Externalities and Cost of Capital. *Management Science*, **60** (9), 2223–2247.

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

Chen, S. and Heese, C. (2021). *Fishing for Good News: Motivated Information Acquisition*. CRC TR 224 Discussion Paper Series, University of Bonn and University of Mannheim, Germany.

Chen, T.-H., Ho, R.-J. and Liu, Y.-W. (2019). Investor personality predicts investment performance? A statistics and machine learning model investigation. *Computers in Human Behavior*, **101**, 409–416.

Coibion, O., Gorodnichenko, Y. and Weber, M. (2022). Monetary Policy Communications and Their Effects on Household Inflation Expectations. *Journal of Political Economy*, **130** (6), 1537–1584.

Commission, E. (2019a). Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee, and the Committee of the Regions: The European Green Deal.

— (2019b). Regulation (EU) 2019/2088 of the European Parliament and of the Council of 27 November 2019 on sustainability-related disclosures in the financial services sector (Text with EEA relevance). Legislative Body: EP, CONSIL.

— (2020a). Regulation (EU) 2020/852 of the European Parliament and of the Council of 18 June 2020 on the establishment of a framework to facilitate sustainable investment, and amending Regulation (EU) 2019/2088 (Text with EEA relevance).

— (2020b). Regulation (EU) 2020/852 of the European Parliament and of the Council of 18 June 2020 on the establishment of a framework to facilitate sustainable investment, and amending Regulation (EU) 2019/2088 (Text with EEA relevance). Legislative Body: CONSIL, EP.

— (2021a). Comission Delegated Regulation (EU) of 21.04.2021 amending Delegated Regulation (EU) 2017/565 as regards the integration of sustainability factors, risks and preferences into certain organisational requirements and operating conditions for investment firms.

— (2021b). FAQ: What is the EU Taxonomy and how will it work in practice?

— (2022a). Commission Delegated Regulation (EU) 2022/1288 of 6 April 2022 supplementing Regulation (EU) 2019/2088 of the European Parliament and of the Council with regard to regulatory technical standards specifying the details of the content and presentation of the information in relation to the principle of do no significant harm, specifying the content, methodologies and presentation of information in relation to sustainability indicators and adverse sustainability impacts, and the content and presentation of the information in relation to the promotion of environmental or social characteristics and sustainable investment objectives in precontractual documents, on websites and in periodic reports.

— (2022b). Directive (EU) 2022/2464 of the European Parliament and of the Council of 14 December 2022 amending Regulation (EU) No 537/2014, Directive 2004/109/EC, Directive 2006/43/EC and Directive 2013/34/EU, as regards corporate sustainability reporting (Text with EEA relevance).

— (2024). Corporate sustainability reporting - European Commission.

Cooper, A. (1999). The Inmates are Running the Asylum. In U. Arend, E. Eberleh and K. Pitschke (eds.), *Software-Ergonomie '99*, vol. 53, Wiesbaden: Vieweg+Teubner Verlag, pp. 17–17, series Title: Berichte des German Chapter of the ACM.

—, Reimann, R., Cronin, D., Noessel, C., Csizmadi, J. and LeMoine, D. (eds.) (2014). *About Face: the essentials of interaction design ; [the completely updated classic on creating delightful user experiences]*. Indianapolis, Ind: Wiley, 4th edn.

Cordes, H., Decke, P. and Schneider, J. C. (2024). The EU Taxonomy in Action: Sustainable Finance Regulation and Investor Preferences.

Cronqvist, H., Siegel, S. and Yu, F. (2015). Value versus growth investing: Why do different investors have different styles? *Journal of Financial Economics*, **117** (2), 333–349.

Crumpler, H. and Grossman, P. J. (2008). An experimental test of warm glow giving. *Journal of Public Economics*, **92** (5-6), 1011–1021.

DAI (2024). Aktionärszahlen 2023 erneut über 12 Millionen.

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

Degryse, H., Di Giuli, A., Sekerci, N. and Stradi, F. (2023). Sustainable Investments: One for the Money, Two for the Show. *SSRN Electronic Journal*.

DeMiguel, V., Gil-Bazo, J., Nogales, F. J. and Santos, A. A. P. (2023). Machine learning and fund characteristics help to select mutual funds with positive alpha. *Journal of Financial Economics*, **150** (3), 1–22.

DESTATIS (2023a). Gehaltsvergleich 2022: Neben dem Beruf ist der Bildungsabschluss entscheidend.

— (2023b). *Sektorale und gesamtwirtschaftliche Vermögensbilanzen*. Tech. rep.

— (2024). Nettoeinkommen Monat bis 2023.

Dimson, E., Karakaş, O. and Li, X. (2015). Active Ownership. *Review of Financial Studies*, **28** (12), 3225–3268.

—, Marsh, P. and Staunton, M. (2020). Divergent ESG Ratings. *The Journal of Portfolio Management*, **47** (1), 75–87.

Ding, C. and He, X. (2004). K-means clustering via principal component analysis. In *Proceedings of the twenty-first international conference on Machine learning*, ICML '04, New York, NY, USA: Association for Computing Machinery, pp. 1–29.

Dong, B., Jiao, P. and Bauer, R. (2022). Is it really not about the money? Individual Investors' Beliefs, Ambiguity Perceptions, and Norm-following Propensities in ESG Investing.

Dou, W. W., Taylor, L. A., Wang, W. and Wang, W. (2021). Dissecting bankruptcy frictions. *Journal of Financial Economics*, **142** (3), 975–1000.

Dumrose, M., Rink, S. and Eckert, J. (2022). Disaggregating confusion? The EU Taxonomy and its relation to ESG rating. *Finance Research Letters*, **48**, 1–8.

Dzyabura, D., El Kihal, S., Hauser, J. R. and Ibragimov, M. (2023). Leveraging the Power of Images in Managing Product Return Rates. *Marketing Science*, **42** (6), 1125–1142.

Døskeland, T. and Pedersen, L. J. T. (2016). Investing with Brain or Heart? A Field Experiment on Responsible Investment. *Management Science*, **62** (6), 1632–1644.

— and — (2021). Does Wealth Matter for Responsible Investment? Experimental Evidence on the Weighing of Financial and Moral Arguments. *Business & Society*, **60** (3), 650–683.

D'Amato, V., D'Ecclesia, R. and Levantesi, S. (2022). ESG score prediction through random forest algorithm. *Computational Management Science*, **19** (2), 347–373.

Eccles, R. G., Ioannou, I. and Serafeim, G. (2014). The Impact of Corporate Sustainability on Organizational Processes and Performance. *Management Science*, **60** (11), 2835–2857.

Ellul, A., Pagano, M. and Scognamiglio, A. (2020). Career Risk and Market Discipline in Asset Management. *The Review of Financial Studies*, **33** (2), 783–828.

Engler, D., Gutsche, G. and Smeets, P. (2023). Why Do Investors Pay Higher Fees for Sustainable Investments? An Experiment in Five European Countries.

Epley, N. and Gilovich, T. (2016). The Mechanics of Motivated Reasoning. *Journal of Economic Perspectives*, **30** (3), 133–140.

ESMA (2022). ESMA publishes final guidelines on MiFID II suitability requirements.

— (2023). *Guidelines on certain aspects of the MiFID II suitability requirements*. Tech. rep.

Eyting, M. (2022). Why do we Discriminate? The Role of Motivated Reasoning.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D. and Sunde, U. (2018). Global Evidence on Economic Preferences*. *The Quarterly Journal of Economics*, **133** (4), 1645–1692.

— and Zimmermann, F. (2013). A Taste for Consistency and Survey Response Behavior. *CESifo Economic Studies*, **59** (1), 181–193.

Famulok, J., Kormanyos, E. and Worring, D. (2023). Do Investors Use Sustainable Assets as Carbon Offsets?

Faradynawati, I. A. A. and Söderberg, I.-L. (2022). Sustainable Investment Preferences among Robo-Advisor Clients. *Sustainability*, **14** (19), 12636.

Fedyk, A. (2024). Front-Page News: The Effect of News Positioning on Financial Markets. *The Journal of Finance*, **79** (1), 5–33.

Filippini, M., Leippold, M. and Wekhof, T. (2024). Sustainable finance literacy and the determinants of sustainable investing. *Journal of Banking & Finance*, **163**, 107167.

Flach, P., Hernández-Orallo, J. and Ferri, C. (2011). A coherent interpretation of AUC as a measure of aggregated classification performance. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ICML'11, Madison, WI, USA: Omnipress, pp. 657–664.

Frey, B. S. and Oberholzer-Gee, F. (1997). The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding- Out. *The American Economic Review*, **87** (4), 746–755.

Friede, G., Busch, T. and Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, **5** (4), 210–233.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T. and Walther, A. (2022). Predictably Unequal? The Effects of Machine Learning on Credit Markets. *The Journal of Finance*, **77** (1), 5–47.

Giglio, S., Maggiori, M., Stroebel, J., Tan, Z., Utkus, S. and Xu, X. (2023). *Four Facts About ESG Beliefs and Investor Portfolios*. Tech. Rep. w31114, National Bureau of Economic Research, Cambridge, MA.

—, —, — and Utkus, S. (2021). Five Facts about Beliefs and Portfolios. *American Economic Review*, **111** (5), 1481–1522.

Gino, F., Norton, M. I. and Weber, R. A. (2016). Motivated Bayesians: Feeling Moral While Acting Egoistically. *Journal of Economic Perspectives*, **30** (3), 189–212.

Gneezy, U., Meier, S. and Rey-Biel, P. (2011). When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives*, **25** (4), 191–210.

Golman, R., Hagmann, D. and Loewenstein, G. (2017). Information Avoidance. *Journal of Economic Literature*, **55** (1), 96–135.

—, Loewenstein, G., Molnar, A. and Saccardo, S. (2022). The Demand for, and Avoidance of, Information. *Management Science*, **68** (9), 6454–6476.

Gramegna, A. and Giudici, P. (2021). SHAP and LIME: An Evaluation of Discriminative Power in Credit Risk. *Frontiers in Artificial Intelligence*, **4**.

Grossman, Z. and van der Weele, J. J. (2017). Self-Image and Willful Ignorance in Social Decisions. *Journal of the European Economic Association*, **15** (1), 173–217.

Guenster, N., Brodback, D., Pouget, S. and Wang, R. (2022). The Valuation of Corporate Social Responsibility: A Willingness to Pay Experiment.

Gutsche, G., Köbrich León, A. and Ziegler, A. (2019). On the relevance of contextual factors for socially responsible investments: an econometric analysis. *Oxford Economic Papers*, **71** (3), 756–776.

—, Wetzel, H. and Ziegler, A. (2023). Determinants of individual sustainable investment behavior - A framed field experiment. *Journal of Economic Behavior & Organization*, **209**, 491–508.

— and ZIEGLER, A. (2019). Which private investors are willing to pay for sustainable investments? Empirical evidence from stated choice experiments. *Journal of Banking & Finance*, **102**, 193–214.

GÉRON, A. (2018). *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*. Beijing Boston Farnham Sebastopol Tokyo: O'Reilly, first edition, fifth release edn.

HAALAND, I., ROTH, C. and WOHLFART, J. (2023). Designing Information Provision Experiments. *Journal of Economic Literature*, **61** (1), 3–40.

HABER, S., KEPLER, J. D., LARCKER, D. F., SERU, A. and TAYAN, B. (2022). ESG Investing: What Shareholders Do Fund Managers Represent?

HAND, D. J. (2009). Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Machine Learning*, **77** (1), 103–123.

HARRISON, G. W. (2006). Experimental Evidence on Alternative Environmental Valuation Methods. *Environmental and Resource Economics*, **34** (1), 125–162.

—, LAU, M. I., RUTSTRÖM, E. E. and SULLIVAN, M. B. (2008). Eliciting Risk and Time Preferences Using Field Experiments: Some Metholodical Issues. In J. P. Carpenter, G. W. Harrison and J. A. List (eds.), *Field experiments in economics*, *Research in Experimental Economics*, vol. 10, Bingley: Emerald, pp. 125–218.

— and RUTSTRÖM, E. E. (2008). Chapter 81 Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods. In C. R. Plott and V. L. Smith (eds.), *Handbook of Experimental Economics Results*, vol. 1, Elsevier, pp. 752–767.

HARTZMARK, S. M. and SUSSMAN, A. B. (2019). Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows. *The Journal of Finance*, **74** (6), 2789–2837.

HASTINGS, J. S., MADRIAN, B. C. and SKIMMYHORN, W. L. (2013). Financial Literacy, Financial Education, and Economic Outcomes. *Annual review of economics*, **5**, 347–373.

HEEB, F., KÖLBEL, J. F., PAETZOLD, F. and ZEISBERGER, S. (2023). Do Investors Care about Impact? *The Review of Financial Studies*, **36** (5), 1737–1787.

HETT, F., GILL, A. and TISCHER, J. (2022). Time Inconsistency and Overdraft Use: Evidence from Transaction Data and Behavioral Measurement Experiments. *SSRN Electronic Journal*.

HONG, H. and KACPERCZYK, M. (2009). The price of sin: The effects of social norms on markets. *Journal of Financial Economics*, **93** (1), 15–36.

HOSMER, D. W., LEMESHOW, S. and STURDIVANT, R. X. (2013). *Applied logistic regression*. Wiley series in probability and statistics, Hoboken, NJ: Wiley, third edition edn.

HÄUBL, G. and TRIFTS, V. (2000). Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids. *Marketing Science*, **19** (1), 4–21.

IFNP (2021). Aufklärungsbedarf beim Thema Nachhaltigkeit: 71 Prozent kennen „ESG-Kriterien“ nicht - FINANCIAL PLANNING News.

IFRS (2024). IFRS - IFRS Sustainability Standards Navigator.

IKOTUN, A. M., EZUGWU, A. E., ABUALIGAH, L., ABUHAIJA, B. and HEMING, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, **622**, 178–210.

JACOBS, B. J., DONKERS, B. and FOK, D. (2016). Model-Based Purchase Predictions for Large Assortments. *Marketing Science*, **35** (3), 389–404.

Jansen, B., Jung, S.-G., Nielsen, L., Guan, K. W. and Salminen, J. (2022). How to Create Personas: Three Persona Creation Methodologies with Implications for Practical Employment. *Pacific Asia Journal of the Association for Information Systems*, **14** (3).

Junkus, J. C. and Berry, T. C. (2010). The demographic profile of socially responsible investors. *Managerial Finance*, **36** (6), 474–481.

Kantar (2023). Beliebteste Geldanlagen der Deutschen bis 2023.

Kempf, A. and Osthoff, P. (2007). The Effect of Socially Responsible Investing on Portfolio Performance. *European Financial Management*, **13** (5), 908–922.

Khandani, A. E., Kim, A. J. and Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, **34** (11), 2767–2787.

Kim, A., Yang, Y., Lessmann, S., Ma, T., Sung, M. C. and Johnson, J. E. V. (2020). Can deep learning predict risky retail investors? A case study in financial risk behavior forecasting. *European Journal of Operational Research*, **283** (1), 217–234.

Kirchler, M., Lindner, F. and Weitzel, U. (2018). Rankings and Risk-Taking in the Finance Industry. *The Journal of Finance*, **73** (5), 2271–2302.

Kleffel, P. and Muck, M. (2023). Aggregate confusion or inner conflict? An experimental analysis of investors' reaction to greenwashing. *Finance Research Letters*, **53**, 103421.

Kumar, S. J. and Oommen Philip, A. (2022). Achieving Market Segmentation From B2B Insurance Client Data Using RFM & K-Means Algorithm. In *2022 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES)*, vol. 1, pp. 463–469.

Kölbel, J. F., Heeb, F., Paetzold, F. and Busch, T. (2020). Can Sustainable Investing Save the World? Reviewing the Mechanisms of Investor Impact. *Organization & Environment*, **33** (4), 554–574.

Latino, C. (2023). Surfing the Green Wave: What's in a "Green" Name Change? *SSRN Electronic Journal*.

Laudenbach, C., Weber, A. and Wohlfart, J. (2021). Beliefs About the Stock Market and Investment Choices: Evidence from a Field Experiment. *SSRN Electronic Journal*.

Laudi, M., Smeets, P. and Weitzel, U. (2023). Do Financial Advisors Exploit Responsible Investment Preferences? *SSRN Electronic Journal*.

Levitt, S. D. and List, J. A. (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives*, **21** (2), 153–174.

Lindauer, M., Eggensperger, K., Feurer, M., Biedenkapp, A., Deng, D., Benjamins, C., Ruhkopf, T., Sass, R. and Hutter, F. (2022). SMAC3: A Versatile Bayesian Optimization Package for Hyperparameter Optimization. *Journal of Machine Learning Research*, **23** (54), 1–9.

List, J. A. (2001). Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards. *American Economic Review*, **91** (5), 1498–1507.

— and Shogren, J. F. (1998). Calibration of the difference between actual and hypothetical valuations in a field experiment. *Journal of Economic Behavior & Organization*, **37** (2), 193–205.

Lokanan, M. E. and Sharma, K. (2022). Fraud prediction using machine learning: The case of investment advisors in Canada. *Machine Learning with Applications*, **8**, 100269.

Lundberg, S. M. and Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, Curran Associates, Inc., vol. 30.

Lusardi, A. and Mitchell, O. S. (2008). Planning and Financial Literacy: How Do Women Fare? *American Economic Review*, **98** (2), 413–417.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, vol. 5.1, University of California Press, pp. 281–298.

Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance*, (7), 77–91.

Mavruk, T. (2022). Analysis of herding behavior in individual investor portfolios using machine learning algorithms. *Research in International Business and Finance*, **62**, 1–20.

McGinn, J. J. and Kotamraju, N. (2008). Data-driven persona development. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, New York, NY, USA: Association for Computing Machinery, pp. 1521–1524.

Medeiros, M. C., Vasconcelos, G. F. R., Veiga, and Zilberman, E. (2021). Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods. *Journal of Business & Economic Statistics*, **39** (1), 98–119.

Mellström, C. and Johannesson, M. (2008). Crowding Out in Blood Donation: Was Titmuss Right? *Journal of the European Economic Association*, **6** (4), 845–863.

Miaskiewicz, T. and Kozar, K. A. (2011). Personas and user-centered design: How can personas benefit product design processes? *Design Studies*, **32** (5), 417–430.

Molnar, C. (2022). *Interpretable machine learning: a guide for making black box models explainable*. Munich, Germany: Christoph Molnar, second edition edn.

Momsen, K. and Ohndorf, M. (2020). When do people exploit moral wiggle room? An experimental analysis of information avoidance in a market setup. *Ecological Economics*, **169**, 106479.

Montagnoli, A. and Taylor, K. (2024). Who Cares About Investing Responsibly? Attitudes and Financial Decisions.

Morningstar (2024a). Global Sustainable Fund Flows: Q1 2024 in Review.

— (2024b). Global Sustainable Fund Flows: Q4 2023 in Review.

Muschelli, J. (2020). ROC and AUC with a Binary Predictor: a Potentially Misleading Metric. *Journal of Classification*, **37** (3), 696–708.

Nielsen, L. (2019). *Personas - User Focused Design*. Springer.

Paetzold, F. and Busch, T. (2014). Unleashing the Powerful Few: Sustainable Investing Behaviour of Wealthy Private Investors. *Organization & Environment*, **27** (4), 347–367.

Pastor, L., Stambaugh, R. F. and Taylor, L. A. (2021). Dissecting Green Returns. *SSRN Electronic Journal*.

Patton, A. J. and Weller, B. M. (2022). Risk Price Variation: The Missing Half of Empirical Asset Pricing. *The Review of Financial Studies*, **35** (11), 5127–5184.

Pedersen, L. H., Fitzgibbons, S. and Pomorski, L. (2021). Responsible investing: The ESG-efficient frontier. *Journal of Financial Economics*, **142** (2), 572–597.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, **12** (85), 2825–2830.

Peer, E., Rothschild, D., Gordon, A., Evernden, Z. and Damer, E. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, **54** (4), 1643–1662.

Popescu, I.-S., Hitaj, C. and Benetto, E. (2021). Measuring the sustainability of investment funds: A critical review of methods and frameworks in sustainable finance. *Journal of Cleaner Production*, **314**, 128016.

Pástor, , Stambaugh, R. F. and Taylor, L. A. (2020). Sustainable investing in equilibrium. *Journal of Financial Economics*.

Qian, S., You, H. and Wan, D. (2023). From Pro-environmental Behavior to ESG Fund Investing: Evidence from account-level data in China. *SSRN Electronic Journal*.

Riedl, A. and Smeets, P. (2017). Why Do Investors Hold Socially Responsible Mutual Funds? *The Journal of Finance*, **72** (6), 2505–2550.

Roethlisberger, F. J. and Dickson, W. J. (2003). *Management and the Worker*. Psychology Press.

Rohatgi, S., Kavidayal, P. C. and Singh, K. K. (2022). Saving Patterns and Investment Preferences: Prediction Through Machine Learning Approaches. In B. Unhelker, H. M. Pandey and G. Raj (eds.), *Applications of Artificial Intelligence and Machine Learning*, Singapore: Springer Nature, pp. 61–79.

Rosenberg, M. (1965). *Society and the Adolescent Self-Image*. Princeton University Press.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, **20**, 53–65.

Salminen, J., Guan, K., Jung, S.-G., Chowdhury, S. A. and Jansen, B. J. (2020). A Literature Review of Quantitative Persona Creation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, New York, NY, USA: Association for Computing Machinery, pp. 1–14.

Schisterman, E. F., Faraggi, D., Reiser, B. and Hu, J. (2008). Youden Index and the optimal threshold for markers with mass at zero. *Statistics in Medicine*, **27** (2), 297–315.

Sebestyen, G. S. (1962). *Decision-making Processes in Pattern Recognition*. Macmillan.

Shapley, L. S. (1952). A Value for N-Person Games. *Contributions to the Theory of Games*, **AM-28** (vol II), 307–317.

Sharot, T. and Sunstein, C. R. (2020). How people decide what they want to know. *Nature Human Behaviour*, **4** (1), 14–19.

Shimokawa, T., Suzuki, K., Misawa, T. and Okano, Y. (2009). Predicting investment behavior: An augmented reinforcement learning model. *Neurocomputing*, **72** (16), 3447–3461.

Silva, T. C., Tabak, B. M. and Ferreira, I. M. (2019). Modeling Investor Behavior Using Machine Learning: Mean-Reversion and Momentum Trading Strategies. *Complexity*, **2019**, 1–14.

Sinaga, K. P. and Yang, M.-S. (2020). Unsupervised K-Means Clustering Algorithm. *IEEE Access*, **8**, 80716–80727.

Statista (2021). Studenten in Deutschland - Nettoeinkommen 2021.

Suisse, C. (2023). Global Wealth Report 2023.

Sutter, M., Kocher, M. G., Glätzle-Rützler, D. and Trautmann, S. T. (2013). Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior. *American Economic Review*, **103** (1), 510–531.

Taleb, N. N. (2018). *Skin in the Game: Hidden Asymmetries in Daily Life*. Random House Publishing Group.

Tonin, M. and Vlassopoulos, M. (2010). Disentangling the sources of pro-socially motivated effort: A field experiment. *Journal of Public Economics*, **94** (11-12), 1086–1092.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, Mass: Addison-Wesley Pub. Co.

Umer, H., Kurosaki, T. and Iwasaki, I. (2022). Unearned Endowment and Charity Recipient Lead to Higher Donations: A Meta-Analysis of the Dictator Game Lab Experiments. *Journal of Behavioral and Experimental Economics*, **97**, 1–15.

van Rooij, M., Lusardi, A. and Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial Economics*, **101** (2), 449–472.

Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, H. and Deng, S.-H. (2019). Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimizationb. *Journal of Electronic Science and Technology*, **17** (1), 26–40.

Xu, S., Qiao, X., Zhu, L., Zhang, Y., Xue, C. and Li, L. (2016). Reviews on Determining the Number of Clusters. *Applied Mathematics & Information Sciences*, **10** (4), 1493–1512.

Yang, T. and Ying, Y. (2022). AUC Maximization in the Era of Big Data and AI: A Survey. *ACM Computing Surveys*, **55** (8), 1–37.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, **3** (1), 32–35.

Yu, S.-S., Chu, S.-W., Wang, C.-M., Chan, Y.-K. and Chang, T.-C. (2018). Two improved k-means algorithms. *Applied Soft Computing*, **68**, 747–755.

Curriculum Vitae

04.12.1992	born in Schramberg, Germany
2012	Abitur Gymnasium Schramberg, Germany
09/2013 - 07/2016	Bachelor of Science Economics University of Mannheim, Germany Thesis Supervisor: Prof. Dr. Hans Peter Grüner
10/2016 - 08/2018	Master of Science Management Johannes Gutenberg University Mainz, Germany Thesis Supervisor: Prof. Dr. Andrej Gill
09/2018 - 07/2024	Doctoral Candidate at the Chair of Corporate Finance Johannes Gutenberg University Mainz, Germany Thesis Supervisor: Prof. Dr. Andrej Gill