Does the willingness to pay for sustainable investments differ between stated and incentivized choice experiments?

January 2022

Abstract

By comparing individual decisions in pre-registered stated and incentivized choice experiments on real-world financial products, this paper empirically examines possible hypothetical biases in the estimated preferences for sustainable investments. The incentivized choice experiment was based on an incentive compatible probabilistic scheme, where participants were randomly selected and the chosen investment products were actually bought at the capital market. Based on representative data for overall more than 2,100 individual investors in Germany and France, our econometric analysis with mixed logit models reveals that, in contrast to previous studies, the willingness to pay for sustainable investments is on average not significantly stronger in the hypothetical than in the incentivized treatment. These estimation results are robust to several explanations of hypothetical bias such as social desirability, strategic motives, choice certainty, and perceived winning probability, as well as experimental design choices and country context. In addition, we find that the estimated effects of several individual characteristics on the preferences for investment products are qualitatively and quantitatively similar in the stated and incentivized settings and also very similar to the corresponding estimation results in previous studies. Our empirical analysis thus sheds light on the reliability of previous stated choice experiments and provide guidance for future choice experiments. Furthermore, our estimation results improve the understanding of individual investment decisions, which is crucial from a policy perspective given that individual investors play an important role in financing the transition to a sustainable economy.

JEL classification:

G11, Q56, A13, C25

Keywords:

Sustainable investments, stated choice experiments, real choice experiments, hypothetical bias, mixed logit models, willingness to pay

1. Introduction

Choice experiments considering hypothetical decision scenarios have become an important tool to elicit individual preferences in many economic contexts (e.g., Johnston et al., 2017). Recently, this elicitation method has also found its way into the study of individual financial decisions (e.g., Benartzi and Thaler, 2007), and particularly individual sustainable investing (e.g., Hartzmark and Sussman, 2019). Though, it is well-known that people may choose differently in incentivized and hypothetical settings. The extent of those differences strongly depends on the type of product, elicitation method, and sample considered (for meta-studies, see e.g., List and Gallet, 2001; Murphy et al., 2005; Schmidt and Bijmolt, 2020). Substantial differences especially arise in contexts where goods with normative attributes are considered because people might derive utility already when they state to contribute to a "good" cause, even though they might not necessarily do what they stated to do (e.g., Andreoni, 1990; Johansson and Svedsäter, 2012). Thus, choice experiments that use hypothetical scenarios to elicit individual preferences for sustainable investments might be especially prone to hypothetical bias. Insights into the extent to which hypothetical bias exists in such investment experiments and the extent to which it varies across different countries and samples is thus of enormous importance.

There is, to the best of our knowledge, no comprehensive theory that explains why people choose differently in hypothetical and incentivized settings. Some explanations for disparities between preferences in hypothetical and incentivized settings are that respondents might think that their choices do not have any consequences in real life, and that they therefore give socially desirable answers based on guesses what other people want to hear rather than revealing their own preferences (e.g., Lusk and Norwood, 2009). Other explanations refer to strategic response behavior. For instance, respondents may want to affect the price or provision of the goods in question (e.g., Lusk, McLaughlin, and Jaeger, 2007). Alternatively, experimental designs could lack incentives to induce respondents to answer truthfully (e.g., Zawojska and Czajkowski, 2017) or respondents are uncertain about their decisions (e.g., Loomis 2014), for example, due to choice task complexity, or insufficient knowledge about and familiarity with the goods in question (e.g., Sanjuán-López and Resano-Ezcaray, 2020). Significant differences between hypothetical and incentivized choices are especially observed in contexts where the available alternatives can be described by normative attributes that, for example, refer to the ecological or social consequences associated with the choice of a specific alternative (e.g., Andreoni, 1990; Johansson and Svedsäter, 2012). Examples exist in the area of

willingness to pay for the support of nature conservation projects (e.g., Brown et al., 1996; Cummings and Taylor, 1999)[1], animal welfare projects (e.g., Johansson and Svedsäter, 2012)[2], or the consumption of environmentally friendly, regional, organic or climate-friendly food (e.g., De-Magistris, Gracia, and Nagye, 2013; Moser, Raffaelli, and Notaro, 2014; Menapace and Raffaeli, 2020), where the willingness to pay for normative attributes is usually higher in hypothetical contexts than in incentivized contexts[3].

Stated choice experiments are increasingly applied in the field of sustainable investments (e.g., Gutsche and Ziegler, 2019; Hartzmark and Sussman, 2019; Lagerkvist et al., 2020), where investment products, for example, consider different normative (e.g., ecological or social) screens to avoid investments that are considered unsustainable. Given the broad empirical evidence from studies that examine goods with normative attributes, results from these surveys and experiments are therefore particularly suspected to be affected by hypothetical bias (e.g., Bauer, Ruof, and Smeets, 2021). We thus set out to investigate whether this is indeed the case and ask three key questions: Do respondents have weaker preferences for sustainable investments in incentivized settings than in purely hypothetical settings? To what extent do reasons identified in previous studies (e.g., social desirability or knowledge about the good) drive differences between incentivized and hypothesized decisions? Are the determinants of sustainable investments identified in past studies the same in incentivized and hypothetical settings?

The analysis is based on pre-registered (no reference due to blind review) choice experiments among representative samples of 2,153 individual investors from France and Germany collected from May to July 2021. Individual investors are defined as financial decision makers in their household who have made experiences with or gained sufficient knowledge of financial products with variable returns (e.g., bonds, bond funds, stocks, equity funds, or more

[1] For example, Brown et al. (1996) find that more than 50% of participants from Wisconsin in the United States in a field study state to be willing to pay for a national park in a hypothetical setting, but only 17% do so in a revealed preference setting. Cummings and Taylor (1999) find significant differences between hypothetical and real contributions to Nature Conservancy projects among Georgia State University Students from the United States.

[2] Johansson and Svedsäter (2012) find a much higher hypothetical than real-money willingness to contribute to a World Wildlife Fund project among 160 students from Gothenburg in Sweden.

[3] De-Magistris, Gracia, and Nagye (2013) report a higher willingness to pay for organic and regional food in a hypothetical compared to an incentivized setting among a sample of residents from Zaragoza in Spain. Moser, Raffaelli, and Notaro (2014) show that there is a considerable hypothetical bias in the willingness to pay for organic and climate friendly food among 288 supermarket customers from Trentino in Italy, and Menapace and Raffaeli (2020) find evidence that the hypothetical willingness to pay for environmentally and socially friendly food exceeds the real willingness to pay among a sample of 498 customers from an organic grocery chain in Italy.

complex assets). Thus, the analysis relies on a large representative sample of participants who are familiar with the good in question. In addition, the setting allows to analyze whether the results generalize to other countries (e.g., Ehmke, Lusk, and List, 2008). In the choice experiment, participants are endowed with €500 and choose six times among four real bond funds that are actually traded on the market and which differ with respect to their strength of sustainability. The experiment consists of four treatments that only differ concerning the information participants receive about incentives and the provision of a safe option (bank account) that is only available in two of the four treatments.

The analysis using mixed logit models in willingness to pay space with correlated random coefficients to analyze willingness to pay for the strength of sustainability reveals that respondents' willingness to pay for sustainability is not significantly higher in the hypothetical compared to the incentivized treatment. This finding holds in both countries considered, France and Germany, and independent of the provision of a safe option. In contrast to what the literature predicts, willingness to pay for sustainability is even higher in the incentivized treatment without safe option in Germany. As participants in the incentivized treatments spend more time on the experiment, read the sustainability attribute description more often, and are more certain about their choices than in the hypothetical treatments, they might think more about the real-world consequences of their choice when it can entail an immediate real-world impact. Those findings are robust to alternative explanations of hypothetical bias such as strategic motives, choice certainty, and perceived winning probability.

Moreover, preferences for sustainable investments in the hypothetical treatments are not significantly stronger for participants with stronger social desirability motives compared to the incentivized treatments. However, in all treatments preferences for sustainable investments are similarly and slightly stronger for participants with stronger social desirability motives. Lastly, the determinants of sustainable investments are found to be qualitatively and quantitatively similar in the incentivized and hypothetical treatments. Those determinants are also similar to the determinants identified in past studies considering stated choice experiments with hypothetical investment decisions (e.g., Gutsche and Ziegler, 2019).

This paper makes four main contributions. First, it extends the literature on differences between incentivized and hypothetical choices in settings that involve goods with normative attributes and consistently find that preferences for sustainable investments are, in contrast to literature-based expectations, not significantly stronger in the hypothetical treatments compared to the incentivized treatments. Second, it provides, to the best of our knowledge, a novel

application of choice modelling by using tradable real-world investment products as basis for the experimental design, which is then combined with a validated incentive mechanism from behavioral economics (e.g., Charness, Gneezy, and Halladay, 2016). Third, the results provide evidence for the validity of past stated choice experiments in the fields of experimental and sustainable finance, show that the determinants of sustainable investments are the same in incentivized and hypothetical settings, and provide guidance for the design of future experiments in those fields. Fourth, using representative samples from two different countries enhances the generalizability of the results, and the study thus contributes to the scarce literature on hypothetical bias country differences (e.g., Ehmke, Lusk, and List, 2008) especially in the willingness to pay for normative attributes. Taken together, those results improve the understanding of individual investor behavior, which is crucial from a policy perspective given that individual investors play an important role in financing the transition to a low-carbon economy (European Commission, 2021).

The remainder of this paper is structured as follows: Section 2 describes the survey including the investment experiment and defines all relevant variables for the empirical analysis. Section 3 presents descriptive statistics, discusses the main estimation results, and includes several robustness checks. Section 4 concludes.

2. Data and variables

The analysis is based on pre-registered (no reference due to blind review) stated and incentivized choice experiments among representative samples of 2,153 individual investors from France and Germany. The survey was carried out in collaboration with the professional market research institute Psyma+Consulting GmbH from May to July 2021. Following earlier studies (e.g., Gutsche and Ziegler, 2019), the survey's target group consists of household's financial decision makers aged 18 or older who have made experiences with or gained sufficient knowledge of financial products with variable returns (e.g., bonds, bond funds, stocks, equity funds, or more complex assets). Such a broad group of individuals is considered as restricting the focus on specific groups of investors (e.g., investors who already own sustainable investment products) might not reflect investment behavior market wide (e.g., Hartzmark and Sussman, 2019).

The survey consists of eight different parts: Part 1 contains questions that allow to screen-out respondents who do not correspond to the target group (e.g., in terms of age, gender, main place of residence), i.e. individuals younger than 18 years of age and/or individuals that did

not make experiences with (or have gained sufficient knowledge of) financial products with variable returns. This part also includes further questions about respondents' current forms of investments. Part 2 comprises several general questions on the respondents' investment and consumption behavior. Part 3 aims to capture a variety of individual characteristics such as economic preferences, personality traits, or personal attitudes. Part 4 consists of the incentivized investment choice experiment which will be described in more detail in the next section (Section 2.1 "Experimental design"). Part 5 captures further background information on the respondents' sustainable investment behavior and knowledge. It particularly contains measures to capture individual financial performance perceptions concerning sustainable investments. Part 6 contains questions regarding low-carbon infrastructure. Part 7 contains questions on financial literacy and cognitive reflection. Finally, part 8 comprises further questions on the respondents' socio-demographic background.

2.1 Experimental design

The empirical analysis in this study is based on an investment choice experiment. At first, participants are randomly allocated to four different treatments as shown in Table 2. In the first group, participants make incentivized choices among real bond funds where no safe option is available. They obtain the same basic information on the setting, i.e. they get a short explanation of the bond funds[4], in which financial products funds can invest (corporate bonds, public bonds, cash, and other derivatives), that the funds accumulate earnings, are traded in Euros, that they have similar risk-return profiles, and that the largest share of the funds' portfolio invests in corporate bonds. As shown in Table 2, the bond funds differ with respect to *Strength of sustainability*, *Annual returns in the past two years*, *Share of bond issuers from the EU*, and *Fees*. In line with Hartzmark and Sussman (2019), the fund's strength of sustainability is based on the Morningstar Sustainability Rating which can range between one and five globes[5].

< insert Table 1 here >

[4] Bond funds were chosen for this experiment as there are relatively low barriers for investing in bond funds for individual investors with respect to liquidity, minimum investment amount, and accessibility compared to, for example, single bonds. They can also directly finance sustainable projects by buying green bonds. In addition, the bond funds in our investment universe vary considerably in terms of sustainability, past performance, and fees, which is advantageous for the statistical efficiency of the experimental design.

[5] See Morningstar (2022) for a detailed description of the Morningstar sustainability rating methodology.

After learning about the setting and attributes, participants are endowed with €500 and get the information that for 10 randomly selected participants of the study, one of their choices of bonds funds is randomly selected, and that the corresponding bond fund is actually bought for them with their endowment. They are informed that their payoff then is the value of the bond fund one year in the future (July 2022 as the study was completed in July 2021), and receive examples of how their payoff is calculated when the value of the fund decreases to €450 or increases to €550 Euros, that is, they receive €450 or €550 minus the fees for the respective fund. They are also guaranteed that all this information is true and will be implemented, and that they are completely free in their decisions. Then, they are asked to choose six times among four randomly drawn alternatives out of a universe of 16 real bonds funds (and a bank account in two out of four treatments as additional fifth alternative). To provide participants with realistically high one-time investment amounts and to reduce administrative complexity, we follow earlier experimental studies analyzing individual investment behavior and only a pay randomly chosen subset of participants (e.g. Kirchler, Lindner, and Weitzel, 2018). Results from various review studies show that such an approach leads to only minor differences, if any, compared to the case where all participants are paid (e.g., Charness, Gneezy, and Halladay, 2016; Clot, Grolleau, and Ibanez, 2018). After the survey, we did indeed invest real money according to the investment decisions. Corresponding to the approach by Falk et al. (2018), endowments were scaled by median household in-come in both countries, and rounded to the next multiple of 100 to facilitate calculations. The reference endowment was €500 in Germany, and scaling resulted in the same €500 endowment for France. The data resulting from this experiment can then be analyzed using mixed logit models, which allow to estimate the willingness to pay for the attributes of the funds, especially the strength of sustainability.

Bond funds were considered for the inclusion into the investment universe when they could be bought by individual investors in Germany either at a stock exchange or directly from the provider of the bond fund. The minimum investment amount could not exceed €250. In addition, the sixteen bond funds were selected such that the attribute values of the different attributes were almost uncorrelated across alternatives. Concerning the strength of sustainability and in line with Hartzmark and Sussman (2019), four bond funds were considered for one, four bond funds for two, four bond funds for four, and four bond funds for five globes, respectively. In addition, we only considered actively managed bond funds that invest a majority of their assets in a portfolio of corporate and public bonds and which may also include other

positions such as cash and other financial products (e.g. derivatives). All the bond funds considered reinvest income in the fund, are traded in euros (€), invest the majority of their portfolio in corporate bonds and have very similar risk and return profiles, that is, two or three according to the German key investor information document[6]. The corresponding funds and attribute values are shown in Table x in the appendix.

Each choice set was constructed by a random draw of four out of sixteen bond funds from the underlying universe. To prevent respondents from getting additional information, for example via financial information providers, the names of the actual bond funds are hidden, and respondents thus can only consider the information provided in the experiment for making their choices. We therefore avoid that participants' familiarity with certain fund providers' names, origin, or other products impacts our results. In each choice set, participants were able to click on question marks next to the attribute names to read the explanation of the attribute again. Figure 1 shows a screenshot of an exemplary choice set. A detailed description and further screenshots of the experiment are shown in the appendix.

The experimental setting is nearly identical for all treatment groups depicted in Table 1. Treatments only differ with respect to the information participants receive concerning incentives and whether a safe option is provided. In the second treatment (II), participants are asked to choose as if they really were to invest €500 and would receive the payoff in July 2022. They also receive examples of how their payoff would be calculated when the value of the fund decreases to €450 or increases to €550, i.e. that they would receive €450 or €550 Euros minus the fees for the respective fund. Apart from the hypothetical instead of incentivized choice, the setting is the same as in the first treatment (I). In the third treatment (III), the setting is again the same as in the first treatment with incentivized choices except that participants are informed that they can also leave their endowment on a bank account and that they will receive €500 in any case in July 2022 when they choose the bank account instead of a bond fund. The bank account option is also available in the fourth treatment (IV), but the choice is hypothetical. Other than that, the setting is again exactly the same as in the other treatments.

2.2 Experiment variables

To capture respondents' investment decisions, we construct the variable *Choice* that takes the value of one for the alternative chosen by the respondent in the respective choice set, and zero

[6] The risk and return profile is based on the synthetic risk and reward indicator. See Committee of European Securities Regulators (2010) for a detailed description of the methodology.

for all non-chosen alternatives. Several attribute-specific variables capture the characteristics of the alternatives offered in the different choice sets. The variable *Strength of sustainability* takes the values of one ("very low"), two ("rather low"), three ("rather high"), and four ("very high") that correspond to the attribute levels for the strength of sustainability for the respective alternative. The variable *Annual returns in the past two years* denotes a bond fund's average return in the years 2019 and 2020 and thus ranges between -0.02% and 12.75%. The variable *Share of issuers of bonds from the European Union* indicates the percentage share of issuers of bonds held by the fund that are from countries of the European Union. The variable thus ranges between 0.00% and 83.23%. Finally, the variable *Fees* captures the amount of fees charged on each alternative and ranges between 0.00% and 6.67%. In addition, the variable *Bank account* takes the value of one to indicate the safe option, if there is any, and zero otherwise. The provision of a safe ("none") option is meant to increase the realism of the choice experiment as investors also have the possibility not to invest and has been found to affect willingness to pay estimates in some contexts (e.g., Penn, Hu, and Cox, 2019). Past studies also show that the absence or presence of incentives affect the frequency of "none"-choices (e.g., Lusk and Schroeder, 2004; Mørkbak, Olsen, and Campbell, 2014).

Finally, we construct several auxiliary variables. We create a dummy variable for each treatment group: The variable *Incentivized choice and no safe option* takes the value of one when a participant was assigned to treatment group I (Incentivized choice and no safe option), and zero otherwise. The variable *Hypothetical choice and no safe option* takes the value of one when a participant was assigned to treatment group II, and zero otherwise. The variable *Incentivized choice and safe option* takes the value of one when a participant was assigned to treatment group III, and zero otherwise. Finally, the variable *Hypothetical choice and safe option* takes the value of one when a participant was assigned to treatment group IV, and zero otherwise. We capture the median duration in seconds that participants needed to go from the introduction screen of the experiment to the last choice set by the continuous variable *Duration*. Finally, we capture how much attention investors paid to the strength of sustainability attribute with the variable *Read sustainability description again* that takes the value one when a respondent clicked on the attribute description of the strength of sustainability in one of the choice sets, and zero otherwise.

2.3 Survey variables

Following previous studies, we additionally capture a wide variety of potential factors that could drive differences between incentivized and hypothetical choices. After each choice set,

we asked respondents how certain they were with respect to their decision. This was done to examine the extent to which respondents felt to have been able to cope with the overall experimental setting (e.g., Loomis, 2014), as they could choose more randomly when they are uncertain about their choices. Respondents could answer on a scale from one to five, where one means "very uncertain" and five means "very certain." The responses to this question are coded as dummy variable *Choice certainty* that takes the value of one when a respondent was rather or very certain about the choice, and zero otherwise. After completion of the choice experiment, we also asked respondents to describe in one sentence what they think the study will be used for. If participants believe that their answers to the survey could influence policies regarding sustainable investments, they could, for example, strategically exaggerate or understate their willingness to pay to signal an increase or decrease in demand (e.g., Mitani and Flores, 2014). The corresponding variable *Strategic behavior* is a dummy variable that takes the of value one when respondents referred to the assessment of preferences for ecological, climate-friendly, or other sustainable investments as possible objective of the study.

To capture social desirability motives, we included the following six items from the Balanced Inventory of Desirable Responding (BIDR) developed by Paulhus (1984, 1991) in a random order: a) "My first impression of people usually turns out to be right," b) "I am very confident of my judgement," c) "I always know why I like things," d) "I have received too much change from a salesperson without telling him or her," e) "I am always honest towards other people," and f) "There have been occasions when I have taken advantage of someone." Items a) to c) capture self-deceptive enhancement and items d) to f) impression management. Individuals who rank high in self-deceptive enhancement are unaware of their overly positive self-presentation, while individuals who rank high in impression management recognize that they are self-enhancing. Respondents could rate their agreement with each statement on a five-point Likert scale ranging from "not at all" (Likert scale one) to "completely" (Likert scale five). After reversing the negative statements d) and f), we give one point for every four or five. The variables *Self-Deceptive Enhancement* and *Impression Management* are the sum of the points for the corresponding three items. Thus, both variables can take values from zero to three.

Investors derive utility from investing in assets that generate a positive, measurable social and environmental impact alongside a financial return (e.g., Barber, Morse, and Yasuda, 2021). Therefore, respondents are asked on a scale from one to five, where one means "not at all" and five means "very strongly", to what extent they agree with the statement "In your investment decisions, a bond fund with a high or very high strength of sustainability was available

at least once. In your opinion, how strongly do such investments contribute to sustainable development?". The responses to this question are coded as dummy variable *Impact* that takes the value of one when a respondent indicates that bond funds rather or very strongly contribute to a sustainable development, and zero otherwise. Moreover, we acknowledge studies that show a relationship between knowledge about and familiarity with the goods in question and hypothetical bias (e.g., Sanjuán-López and Resano-Ezcaray, 2020). Consequently, we implemented a widely accepted objective measure of financial literacy designed by Lusardi and Mitchell (2008). This measure is based on three quiz questions about interest rates, inflation, and risk diversification, respectively, and thus aims to capture an individual's fundamental economic and financial knowledge. Accordingly, the number of correct answers can range between zero and three, and the corresponding dummy variable *Financial literacy* takes the value of one when a respondent answers the same or more correct answers than the sample median for the respective country, and zero otherwise. Lastly, to capture familiarity with sustainable investments, the variable *Knows sustainable investments* takes the value of one when a respondent indicated to have heard of sustainable investments before the study, and zero otherwise.

Respondents in the incentivized treatments are additionally asked on a scale from one to five, where one means "very unlikely" and five means "very likely," how likely they think it is that they will receive the payoff from the experiment to examine whether choices differ with the perceived winning probability in the experiment. The responses to this question are coded as dummy variable *Perceived winning probability* that takes the value of one when a respondent thought he or she was rather or very likely to receive the payoff, and zero otherwise.

Finally, to examine the extent to which the identified determinants of individual sustainable investment behavior differ based on hypothesized and incentivized experiments, we create variables similar to those considered by Gutsche and Ziegler (2019). Willingness to pay for sustainable investment products might be influenced by pecuniary factors such as a lower expected risk of sustainable investments. Therefore, we asked respondents to what extent they agree with the statement "Sustainable investments are riskier than conventional investments" on a symmetric scale with the five ordered response categories "strongly disagree," "rather disagree," "neither agree nor disagree," "rather agree," or "completely agree.". On this basis, the dummy variable *High perceived risk* is created that takes the value of one if a respondent indicated one of the two highest categories, and zero otherwise. Apart from pecuniary factors,

past empirical analyses have underlined the importance of non-pecuniary psychological motives, values, and norms for sustainable investment decisions. One motive that is often important for sustainable behavior is warm glow, which can be described as a good feeling through the act of giving (e.g., Andreoni, 1990). Such feelings can lead to psychological benefits and thus higher utility levels from acting sustainably. The corresponding dummy variable *Warm glow* takes the value of one if respondents agreed rather strongly or totally to at least one the statements "It makes me feel good when I behave sustainably "or "I feel responsible for sustainable development and want to contribute to it through my behavior". To capture potential effects of social norms (e.g., Gutsche, Köbrich, and Ziegler, 2019), the dummy variable *Expectation social environment* takes the value of one if a respondent agreed rather strongly or totally with the statement "My social environment (e.g. family, friends, colleagues) expects me to behave in a sustainable manner". In addition to these two non-pecuniary motives, we also consider indicators for sociability (e.g., Gutsche, Köbrich-Leon, and Ziegler, 2019) and political identification (e.g., Gutsche and Ziegler, 2019). The corresponding dummy variable *Volunteering* takes the value of one when individuals report to be engaged in volunteering activities, and zero otherwise. Concerning political identification, past studies find evidence that a left-wing (e.g., Gutsche, Köbrich-Leon, and Ziegler, 2019) and ecological policy identification (e.g., Gutsche and Ziegler, 2019) are correlated with sustainable investments. Following Ziegler (2017, 2019), we therefore measure individual political identification using the dummy variables *Social policy identification* and *Ecological policy identification*. The variable *Social policy identification* takes the value one if the respondent rather or totally agreed with the statement "I identify myself with socially oriented politics" and the variable *Ecological policy identification* takes the value one if the respondent rather or totally agreed with the statement "I identify myself with ecologically oriented politics."

In addition, we capture usual socio-demographic characteristics. The dummy variable *Female* takes the value of one if a respondent is a woman, and zero otherwise. The variable *Age* denotes a respondent's age in years. The dummy variable *High education* takes the value of one if a respondent at least has a university entrance qualification according to level six of the International Standard Classification for Education (Eurostat, 2022). The dummy variable *Living together or married* takes the value of one for these two marital statuses, and zero otherwise. Finally, the dummy variable *Western Germany* takes the value of one if a respondent lives in one of the Western German federal states excluding Berlin.

2.4 Sample characteristics and randomization

Table 2 reports the sample characteristics for all four treatment groups and both countries. This overview reveals that respondents tend to be male, are older, and highly educated as we consider household financial decision makers with at least some investment knowledge. That was to be expected since previous studies find similar individual investor characteristics (e.g., Riedl and Smeets, 2017; Choi and Robertson, 2020). In terms of randomized allocation of respondents to the four different treatments, we see mostly no significant differences in individual characteristics between the treatments within each country. This implies that the randomization was successful[7].

Concerning the experimental variables, respondents in the incentivized treatment without safe option are significantly more certain about their choices compared to those in the hypothetical treatment without safe option in both countries. The same pattern arises when a safe option is introduced in Germany, but the difference is not significant. Respondents spend more time on the experiment in the incentivized settings in both countries. They are also more likely to read the description of the sustainability attribute description again in the incentivized treatments, but the difference is significant only for France. There are also no significant differences for strategic behavior found for both countries. The general impact of bond funds on sustainable development is not assessed significantly differently between treatments in Germany. In contrast, respondents in the incentivized treatment without safe option in France are significantly more likely to assess the impact to be rather or very strong compared to the hypothetical treatment. The interpretation of those findings is provided in chapter 3.2.

3. Econometric analysis

3.1 Econometric approach

In the following, we briefly depict the econometric approach used to analyze the data derived from the different versions of our discrete choice experiment. Accordingly, in the two treatments without safe option, each respondent i chooses $M = 6$ times among $J = 4$ bond funds.

[7] Table A.1 in the appendix reports the differences in the means between the four treatment groups within each country and the significance levels based on the mean comparison z-test. For 105 comparisons, about one difference is expected to be significantly different from zero at the one percent significance level, about five differences to be significantly different from zero at the five percent significance level, and about ten differences to be significantly different from zero at the ten percent significance level. Table A.1 reveals that no difference is significantly different at the one percent significance level, two differences are significantly different at the five percent significance level, and six differences are significantly different at the ten percent significance level. Therefore, the number of significant differences is lower as statistically expected.

In the other two treatments, each respondent i chooses $M = 6$ times among $J = 5$ alternatives (four bonds funds and a bank account). Moreover, the alternatives in each choice set vary in terms of the attributes captured by the variables *Strength of sustainability*, *Annual returns in the past two years*, *Share of issuers of bonds from the European Union*, and *Fees*. In treatments with a safe option, this additional alternative is indicated by the variable *Bank account*. To examine the relevance of these attributes on the choice among the four (or five) available mutually exclusive alternatives in an econometric analysis, we thus consider multinomial discrete choice models, assuming utility functions for each choice alternative. In our case, the utility of respondent i ($i = 1,\ldots, N$) for investment alternative j ($j = 1,\ldots, 4$ *or* $j = 1,\ldots, 5$) in choice set m ($m = 1,\ldots, 6$) is:

$$U_{ijm} = \beta_i' X_{ijm} + \varepsilon_{ijm}$$

An individual's utility U_{ijm} thus depends on the vector $x_{ijm} = (x_{ijm1},\ldots, x_{ijmK})'$ of explanatory variables that are based on the attributes and individual characteristics as well as on the corresponding unknown parameter vectors β_i ($\beta_i = \beta_{i1},\ldots, \beta_{iK})'$. $K = 5$ and $K = 6$ explanatory variables are included in the models without and with including the bank account alternative, respectively.

The values of U_{ijm} cannot be observed and depend on the error terms ε_{ijm}, which summarize all unobserved factors for the choice of an investment alternative. According to random utility maximization theory (e.g., McFadden, 1973), it is assumed that respondents choose an investment alternative in a specific choice set if the utility for this alternative is the largest among the utilities for all four (or five) alternatives. To avoid the well-known problems associated with common multinomial logit models, which have been a workhorse to analyze such econometric problems over many years, we apply mixed logit models (e.g., McFadden and Train, 2000). These models also assume independently and standard (type 1) extreme value distributed error terms ε_{ijm}, but do not require the restrictive independence of irrelevant alternatives assumption. Mixed logit models specifically assume that the parameters $\beta_i = \beta$ ($i = 1,\ldots, N$) of the explanatory variables are continuously distributed across i. All parameters except for interaction terms are assumed to be random. Based on economic theory, utility associated with (negative) fees cannot be negative and is therefore assumed to be log-normally distributed. The remaining random parameters are assumed to follow normal distributions, as respondents might experience positive or negative utility from each of the corresponding attributes. In the present application, β_i follows the multivariate distribution f(β, Σ), where β is a vector of

sample means and Σ is a variance–covariance matrix that allows to account for unobserved taste heterogeneity and correlations between random taste parameters.

A small reformulation of the utility based on separating the vector of explanatory variables x_{ijm} into price and non-price attributes allows to directly interpret the parameter as willingness to pay in terms of fees for the respective attributes (Train and Weeks, 2005):

$$U_{ijm} = \alpha_i (\omega_i' y_{ijm} - p_{ijm}) + \varepsilon_{ijm}$$

where α_i is the parameter for fees, ω_i is a vector of marginal willingness to pay for each of the K-1 non-fees attributes $y_{ijm} = (y_{ijm1}, \ldots, y_{ijmK-1})$, and p_{ijm} is the fees attribute.

In contrast to multinomial logit models, the maximum likelihood estimation of mixed logit models is not feasible since the probabilities for the choice of the four (or five) alternatives (which are components of the loglikelihood function) are characterized by multiple integrals, which cannot be computed with deterministic numerical integration methods. Instead, the probabilities can be approximated by simulation methods. The inclusion of these simulated probabilities leads to the simulated maximum likelihood (SML) estimation (e.g. Revelt and Train, 1998; Train, 2009). For the SML estimation of mixed logit models, we use the R package "Apollo" (Hess and Palma, 2019) and R = 2000 draws based on the Modified Latin Hypercube Sampling algorithm.

3.2 Do respondents have weaker preferences for sustainable investments in incentivized settings than in purely hypothetical settings?

To answer our main research questions, Table 3 reports the willingness to pay estimates for the different attributes of the investment alternatives. Considering the German sample, we see that respondents from the incentivized choice and no safe option group are on average willing to pay 2.35 percentage points higher fees when a funds' strength of sustainability increases by one globe. Respondents whose investment decision was also incentivized but who were also offered a safe option are on average willing to pay 1.95 percentage points higher fees for an increase of the strength of sustainability by one category. The estimated mean willingness to pay for an increase of the strength of sustainability by one category for the French sample amounts to 1.96 percentage points. We thus see very similar average willingness to pay estimates based on incentivized investment decisions regardless of the country considered or the

availability of an additional safe option.[8] These results are in line with previous studies based on stated choice experiments implying that individual investors are willing to sacrifice returns to invest in a sustainable manner (e.g., Gutsche and Ziegler, 2019; Hartzmark and Sussman, 2019; Lagerkvist et al., 2020). In addition, we see strong evidence for unobserved heterogeneity in the willingness to pay for the strength of sustainability based on the corresponding standard deviation parameters shown in Table 3.

To what extent can we observe differences in mean willingness to pay estimates in incentivized and hypothetical settings? Such differences would be absorbed by the interaction term, which we included in all three models. Table 3 shows that this interaction term is statistically significantly different from zero in only one of the three models. Thus, in two of the three approaches, we see no evidence of divergent mean willingness to pay estimates based on incentivized and hypothetical discrete choice experiments. The significant interaction term implies that respondents in the hypothetical treatment group without safe option are willing to pay 0.45 percentage points less for an increase of the strength of sustainability by one category compared to the group with an incentivized choice and no safe option.[9] This finding is in stark contrast to the significantly higher mean willingness to pay in the hypothetical setting that was expected based on past studies on hypothetical bias for goods with normative attributes.

Result 1a: *There is no evidence that the willingness to pay for sustainable investments is lower for incentivized decisions compared to hypothetical decisions***.**

The results in Table 3 are also not in line with previous studies reporting evidence that the direction and degree of hypothetical bias differ across countries (e.g., Ehmke, Lust, and List, 2008). In both Germany and France, we see no evidence of higher average willingness to pay estimates for the sustainability attribute. Therefore, the previous result is independent of the

[8] Concerning all further attributes, Table 3 shows that respondents are on average willing to pay between 0.45 to 0.52 (0.68) percentage points more fees for an increase of annual returns in the past two years by one percentage point in Germany (France). For the share of issuers of bonds from the European Union, respondents are on average willing to pay between 0.01 and 0.02 (0.02) percentage points more fees in Germany (France) for a one percentage-point increase. Moreover, all mean willingness to pay coefficients are significantly different from zero at the 1%-level. In addition, the mean parameter for the safe option is not significantly different from zero and implies that respondents on average have no preference to stick with the safe option. However, the large and significant standard deviation parameter for the safe option implies that high shares of respondents either have strong preferences for or against choosing the safe option. In general, all remaining standard deviation parameters are significantly different from zero and thus indicate strong unobserved heterogeneity in the willingness to pay for the respective attributes.

[9] The latter group has a mean willingness to pay for an increase of the strength of sustainability by one category of 2.35 percentage points, as discussed before.

country context, at least in our setting where we consider the same target populations, methodological approaches, dependent variables, explanatory variables, and time periods in both countries.

Result 1b: *The finding that there is no evidence for a lower willingness to pay for sustainable investments in incentivized compared to the hypothetical investment decisions is not limited to one specific country*.

As outlined in section 2.2, another context factor that has been shown to affect willingness to pay estimates and hypothetical bias is the provision of a status quo (or "none") option (e.g., Lusk and Schroeder, 2004). We note that respondents on average have no preference to stick with the safe option based on the corresponding insignificant mean parameter. In contrast, the corresponding standard deviation parameter implies that high shares of respondents either have strong preferences for or against choosing the safe option. However, we find no evidence for a statistically significant difference between the incentivized and hypothetical treatments when a safe option is provided. The willingness to pay for the strength of sustainability is therefore not significantly stronger in the hypothetical compared to the incentivized treatment when a safe option is provided. Incentivized and hypothetical choices thus are more similar when a safe option is provided, which might be attributed to the higher realism of the choice setting when a safe option exists (e.g., Lusk and Schroeder, 2004).

Result 1c: *There is no evidence for a lower willingness to pay for sustainable investments in incentivized compared to the hypothetical investment decisions when a safe option is included in the experimental design.*

While we find no evidence of higher willingness to pay for sustainable bond funds in hypothetical compared to incentivized choices, as described above, we even unexpectedly see significantly lower willingness to pay in the incentivized setting without safe option compared to its hypothetical counterpart in the German sample (see column 1 in Table 3). In France, the corresponding difference is not statistically significant, but the z-statistic of 1.535 is close to the 10%-significance threshold. How can we explain these differences?

Participants spend more time on the experiment when incentives are present in treatments without safe option. As shown in Table 3, among German respondents, the median time spent from the introduction screen of the experiment to the last choice situation is 265 seconds for the incentivized treatment without safe option compared to 234 seconds for the hypothetical

counterpart. Similarly, French participants spend 259 seconds in the corresponding incentivized treatment, and thus 22 seconds more than in the respective hypothetical setting (227). These time differences between incentivized and hypothetical treatments mainly result from the time spent on the introduction screen where the incentive mechanism is explained with about 50 additional words. The median time spent on the single choice sets is between one to three seconds longer in the incentivized treatments.[10] In contrast, participants roughly spend the same time on reading the attribute descriptions across treatments.

Participants in the incentivized treatments without safe option are also significantly more certain about their choices[11]. In France, respondents are less certain about their choices in the incentivized treatment than in the hypothetical treatment. While respondents from Germany are on average more certain about their choices compared to respondents from France, they are similarly less certain about their choices in the incentivized treatment when no safe option is provided.

Incentives thus lead participants to spend more time on the experiment, to more often read the sustainability attribute description again during the choice experiment when no safe option exists, and to be more certain about their choices.[12] With the exception of the time spent on the experiment, those differences are not found in Germany when a safe option is introduced. Taken together, those findings imply that participants care and think more about the real-world impact of their choices when there is a possibility that the funds are actually bought, and therefore show stronger preferences for sustainability when there is no possibility to leave

[10] For example, the median time spent by participants is 42 (39) seconds in the first choice set and 24 (22) seconds for the second choice set in the incentivized treatment without safe option (hypothetical treatment without safe option). Participants usually read the attribute description again in the first choice sets and have to get used to the setting, which might explain why the spend more time on the first choice set than on the following choice sets. For choice sets three to six, no differences occur for the median time spent on the choice sets between the treatments. The same pattern is present for both safe option treatments.

[11] Table A.2 shows that participants in the settings without safe option are significantly less likely to be rather or very certain about their choices in the hypothetical compared to the incentivized treatments in both countries. In contrast, there is no statistically significant difference between the incentivized treatment without safe option and both safe option treatments in Germany.

[12] 7.11% (4.90%) of participants in the incentivized (hypothetical) treatment without safe option click on the attribute description of the sustainability attribute during the experiment in Germany at least once, but the difference between the incentivized and hypothetical treatment is not statistically significant (p-value of 0.185 in a Pearson Chi Square test). In France, 25.55% (17.77%) of participants in the incentivized (hypothetical) treatment without safe option click on the attribute description of the sustainability attribute at least once during the experiment. This difference is significant on the 1%-level (p-value of 0.006 in a Pearson Chi Square test). There are no significant differences between the share of participants who click on the attribute description of costs in the choice sets in Germany (between 5.93% and 9.02% of respondents). In contrast, participants from France are again significantly more likely to click on the description of the cost attribute in the incentivized compared to the hypothetical treatment without safe option (10.81% vs 5.21%).

the endowment on a bank account. Similarly, the additional time spent on the experiment and the stronger focus on the sustainability attribute also lead to more certain choices. Introducing a safe option in Germany leads those differences to disappear. When participants are not certain about their choice but have to choose a bond fund, they might want to ensure that the investment is at least sustainable as a decision heuristic, while they would have chosen the bank account if it was available.

3.3 To what extent do reasons identified in previous studies drive differences between incentivized and hypothesized decisions?

Having found on average no higher willingness to pay for sustainable investments on the basis of the hypothetical experimental setting than in the incentivized setting, we now consider the extent to which possible differences arise with respect to factors that have been identified as drivers of hypothetical bias in previous studies. In other words, we aim to analyze whether certain subgroups on average have a higher willingness to pay for sustainable bond funds in hypothetical than in incentivized settings, although there is no evidence of an upward hypothetical bias overall. To this end, we first social desirability motives, strategic behavior, choice certainty, and different indicators of financial knowledge. For the corresponding analyses, we consider the comparison between the incentivized and hypothetical choice and no safe option groups in Germany. The overall results are, however, similar when the treatments with safe option in Germany or the hypothetical and incentivized treatment groups in France are considered. The estimation results for those comparisons are available upon request.

Social desirability motives

Some respondents might be especially prone to give socially desirable answers and thus report a higher willingness to pay for normative goods like sustainable investments because they derive utility already when they state to contribute to a “good” cause, even though they might not necessarily do what they stated to do (e.g., Andreoni, 1990; Johansson and Svedsäter, 2012). Therefore, we split our sample along the previously described measures of social desirability, namely self-deceptive enhancement and impression management. We particularly only focus on the subgroup of respondents with a strong tendency to give socially desirable answers. To this end, we mainly consider those respondents whose scores for self-deceptive enhancement and impression management are higher than the corresponding sample medians.

The expectation is that respondents who tend to give socially desirable answers have a higher willingness to pay for sustainable investments is lower when the choice is incentivized than

when the choice is hypothetical. However, Table 4a shows that we do not find evidence for a different pattern for these subgroups compared to our overall findings: the willingness to pay for the strength of sustainability is not smaller in the incentivized than in the hypothetical treatment group, which we find for both subgroups with high or low self-deceptive enhancement. The same is found for both subgroups with high or low impression management. However, the subgroups with high self-deceptive enhancement and impression management have a higher willingness to pay for sustainable investments compared to the subgroups with low self-deceptive enhancement and impression management, independent of whether choices are incentivized or hypothetical. Additional estimation results, which are available upon request, imply that these differences are significant.

Result 2a: The finding that there is no evidence for a lower willingness to pay for sustainable investments in incentivized compared to the hypothetical investment decisions is also found for respondents who tend to give socially desirable answers. However, the willingness to pay for sustainable investments is higher in both the hypothetical and incentivized group for respondents with stronger social desirability motives.

Alternative explanations for hypothetical bias from the literature refer to strategic behavior and choice certainty. Respondents who mention that analyzing preferences for sustainable investments was the goal of the study might want exaggerate or understate their willingness to pay for sustainable investments to signal an increase or decrease in demand and thereby, for example, possibly influence regulations. However, we again find no evidence that the willingness to pay for sustainable investments is lower in incentivized compared to the hypothetical treatment groups, neither for the subgroup that mentioned the goal of the study nor the subgroup who did not.

Respondents could have a stronger hypothetical bias when they are not sure about their choice, for example because the choice is too complex or hard to make. The models on the right of Table 4a reports the estimation results when only those choice sets are considered where respondents stated to either have been rather or very certain about their choice or not. Similar to the findings above, we find no evidence in both subgroups that the willingness to pay for sustainable investments is lower in incentivized compared to the hypothetical treatment groups.

In addition, bond funds in the present experimental setting are characterized by both complex financial and sustainability information, and choosing among the available alternatives can thus be difficult for respondents. Incentivized and hypothetical decisions might thus differ

depending on the extent to which respondents understand the choice situation and possibly choose sustainable investments as a simplifying heuristic to avoid cognitive load (e.g., Boxall, Adamowicz, and Moon, 2008). Therefore, estimation results for additional subgroups based on financial literacy and subjective investment knowledge are shown in Table 4b. Considering the models for respondents who either rank below or equal to or above the sample median for financial literacy, we again find no evidence that the willingness to pay for sustainable investments is lower in the incentivized compared to the hypothetical treatment groups. The same is found for respondents who stated to have known sustainable investments before the survey and respondents who did not have known sustainable investments before. Lastly, investors derive utility from investing in assets that generate a positive, measurable social and environmental impact alongside a financial return (e.g., Barber, Morse, and Yasuda, 2021), and they might differ in their perception of the real-world impact that bond funds can have. For both subgroups who think that either think bond funds can rather or very strongly contribute to a sustainable development or not, we again find no evidence that the willingness to pay for sustainable investments is lower in the incentivized compared to the hypothetical treatment groups. However, we note that the willingness to pay for sustainable investments is considerably higher for the subgroup that thinks bond funds have an impact. Additional estimations, which are available upon request, show that this difference is statistically significant. The willingness to pay for sustainable investments in the incentivized treatment is hence not significantly lower than in the hypothetical treatment, independent of objective financial knowledge, familiarity with sustainable investments, and the perceived real-world impact of bond funds.

Lastly, respondents might perceive the chance that they will be selected for receiving the payoff as small and therefore do not answer as if they would definitely receive a payoff. The right side of Table 4b reports the estimation results for all respondents from the hypothetical group and either those respondents from the incentivized group are considered who thought to have a rather or very high winning probability or those who did not think so. Again, there is no evidence that respondents are willing to pay less for sustainable investments in the incentivized than in the hypothetical group for both a high and low perceived winning probability.

Result 2b: Across a wide range of different subsamples, the willingness to pay for sustainable investments in the hypothetical treatment is never found to be significantly higher than in the incentivized treatments.

3.4 Are the determinants of sustainable investments identified in past studies the same in incentivized and hypothetical settings?

In a third step, we examine the extent to which the determinants of individual preferences for sustainable bond funds differ based on incentivized and hypothetical settings. In doing so, we also automatically address the strong empirical evidence of unobserved heterogeneity with respect to willingness to pay for sustainable bond funds, as indicated by the highly significant standard deviations of the random parameters in Table 5. For this purpose, the models presented in Table 5 comprise the possible determinants of individual sustainable investments defined earlier and also considered by Gutsche and Ziegler (2019). This overview clearly shows that both incentivized and hypothetical approaches lead to very similar estimation results.

Similar to the results reported by Gutsche and Ziegler (2019), feelings of warm glow from sustainable activities and an ecological policy orientation are significantly positively correlated with a higher willingness to pay for an increase of a fund's strength of sustainability. The estimated effects not only have the same signs in the incentivized and hypothetical settings, but are also very similar in terms of their magnitude. The estimated willingness to pay is on average 0.90 percentage points higher for respondents who derive a warm glow from sustainable consumption in the incentivized setting, which is very similar to estimate from the hypothetical setting (0.94 percentage points). Likewise, respondents with an ecological policy orientation are on average willing to pay 0.78 percentage points higher fees for bonds funds with a higher sustainability rating in the incentivized setting, compared to 0.96 percentage points in the hypothetical setting.

Similarities also emerge with regard to the perception of a higher risk of sustainable compared to conventional investments. In both settings, the willingness to pay for a bond fund with a higher sustainability rating decreases significantly by 0.33 percentage points in the incentivized and 0.35 percentage points in the hypothetical setting. Gutsche and Ziegler (2019) also find a negative estimated correlation with respect to this variable, but it is not statistically significant different from zero.

In addition to these similarities between the results based on incentivized and hypothetical choices, we also see small differences in terms of some sociodemographic characteristics such as education, relationship status, and regional origin. The significant effects for these variables were also not observed in this manner in the study by Gutsche and Ziegler (2019). While the

results for relationship status and regional origin have the same sign in both models, the estimates for education point in different directions. Possible reasons for the differences are temporal differences, deviations in the experimental designs, or the slightly different sampling processes[13]. However, we can only speculate about this based on our (cross-sectional) data. Apart from that, the remaining explanatory variables do not seem to significantly affect the willingness to pay for sustainable bond funds in either setting.

Result 3: Determinants of sustainable investments are qualitatively and quantitatively similar in the incentivized and hypothetical treatments. Those determinants are also similar to the determinants identified in past studies.

3.5 Robustness checks

Several robustness checks support the findings reported above. We additionally estimated all models in preference space using Halton draws instead of draws based on the Modified Latin Hypercube Sampling algorithm. Similar to the specification in the willingness to pay space models, we also estimated all models with a lognormal distribution of the random coefficient for fees, normal distributions for the coefficients of the remaining attributes, and fixed coefficients for the interaction terms. Moreover, for all models, we also considered simple conditional logit models using fixed coefficients for all attributes and interactions, and used standard errors clustered at the individual level. The findings in all specifications are qualitatively very similar to the results based on models in willingness to pay space. All corresponding estimation results are available on request.

4. Conclusions

Many previous studies show that people make different choices in hypothetical and incentivized experimental settings, and especially exaggerate their willingness to pay when goods with normative attributes are considered in hypothetical settings but not when real money is at stake. On the basis of data from a large-scale computer-based survey experiment on the choice of real-world bond funds among a representative sample of more than 2100 German and French individual investors, our econometric analysis with mixed logit models in willingness to pay space with correlated coefficients shows that the willingness to pay for sustainability

[13] Gutsche and Ziegler (2019) consider household financial decision makers who at least need to have a savings account, whereas we only consider respondents who have made experiences with or gained sufficient knowledge of financial products with variable returns (e.g., bonds, bond funds, stocks, equity funds, or more complex assets).

is not significantly higher in the hypothetical than in the incentivized treatments in contrast to what the literature predicts. This finding is found for both the German and French subsamples and independent of the provision of a safe option. Willingness to pay for the funds strength of sustainability is even higher in the incentivized treatment without safe option in Germany compared to the hypothetical treatment without safe option. As participants in the incentivized treatments spend more time on the experiment, read the sustainability attribute description more often, and are more certain about their choices than in the hypothetical treatments, they might think more about the real-world consequences of their choice when it can entail an immediate real-world impact.

An additional econometric analysis also reveals that the willingness to pay for sustainable investments in the hypothetical treatments is also not significantly stronger for participants with stronger social desirability motives compared to the incentivized treatments. However, in all treatments preferences are similarly and slightly stronger for participants with stronger social desirability motives. Similarly, the findings are robust to a range of alternative explanations of hypothetical bias such as strategic motives, choice certainty, and perceived winning probability.

Lastly, the determinants of sustainable investments are found to be qualitatively and quantitatively similar in the incentivized and hypothetical treatments, and to be similar to the determinants that have been identified in past studies.

Considering the increasing popularity of stated choice experiments in general and particularly in experimental and sustainable finance, the issue of potential differences between hypothetical and incentivized choices is an important topic. The results of this study therefore bear important implications for the validity of past hypothetical studies and the design of future studies. It is not always possible or desirable to consider real-world data in choice experiments, as researchers might be interested in preferences for new attributes that do not yet exist on the market. For example, the results reported in this study about the determinants of sustainable investments lead to very similar conclusions compared to past studies that used hypothetical financial products with artificial attributes (e.g., Gutsche and Ziegler, 2019). This is true for both the incentivized and hypothetical setting. While it is not possible to always use monetary incentives in choice experiments in finance, for example due to budget constraints or the lack of available real-world financial products, the absence of incentives does not necessarily lead to different conclusions.

The results reported here do, however, not necessarily reflect “true” preferences due to the experimental setting even though monetary incentives are at stake. One interesting avenue for further research would be to compare the results with real transaction data, for example, individual investor data from retail (online) brokers, which would add another degree of realism to the analysis (e.g., Haghani et al., 2021).

While we do not find evidence for hypothetical bias for general sustainability as a normative attribute, that is, a combined measure of environmental, social, and governance criteria, it would be interesting to analyze whether the conclusions would be different for those different aspects of sustainability. For example, choosing funds that particularly consider social criteria could be perceived as more socially desirable compared to a more general sustainability measure and therefore be suspect to stronger hypothetical bias.

The estimation results reported here refer to the specific case of Germany and France in 2021. Therefore, it is an open question whether these results are valid for other time periods and countries. Similar empirical analyses in other non-European industrialized countries are an interesting direction for future research. For example, Japanese individual investors have a stronger focus on pecuniary factors when making sustainable investment decisions, and could therefore react stronger on the presence or absence of monetary incentives.

References

Adamowicz, W., D. Dupont, A. Krupnick, and J. Zhang (2011), Valuation of cancer and microbial disease risk reductions in municipal drinking water: an analysis of risk context using multiple valuation methods, *Journal of Environmental Economics and Management* 61 (2), 213–226.

Andreoni, J. (1990), Impure altruism and donations to public goods: A theory of warm-glow giving, *The Economic Journal* 100 (401), 464-477.

Barber, B. M., A. Morse, and A. Yasuda (2021), Impact investing, *Journal of Financial Economics*, 139 (1), 162-185.

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

Bauer, R., T. Ruof, and P. Smeets (2021), Get real! Individuals prefer more sustainable investments, *The Review of Financial Studies* 34 (8), 3976-4043.

Benartzi, S. and R. Thaler (2007), Heuristics and biases in retirement savings behavior, *Journal of Economic Perspectives*, 21 (3), 81-104.

Boxall, P., W. L Adamowicz, and A. Moon (2009), Complexity in choice experiments: choice of the status quo alternative and implications for welfare measurement, *Australian Journal of Agricultural and Resource Economics*, 53 (4), 503-519.

Brown, T. C., P. A. Champ, R. C. Bishop, and D. W. McCollum (1996), Which response format reveals the truth about donations to a public good?, *Land Economics* 152-166.

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

Choi, J. J. and A. Z. Robertson (2020), What matters to individual investors? Evidence from the horse's mouth, *The Journal of Finance*, 75 (4), 1965-2020.

Clot, S., G. Grolleau, and L. Ibanez (2018), Shall we pay all? An experimental test of random incentivized systems, *Journal of Behavioral and Experimental Economics* 73, 93-98.

Committee of European Securities Regulators (2010), CESR's guidelines on the methodology for the calculation of the synthetic risk and reward indicator in the Key Investor Information Document, https://www.esma.europa.eu/sites/default/files/library/2015/11/10_673.pdf (last accessed January 27, 2022)

Cummings, R. G., and L. O. Taylor (1999), Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method, *American Economic Review* 89 (3), 649-665.

De-Magistris, T., A. Gracia, and R. M. Nayga Jr. (2013), On the use of honesty priming tasks to mitigate hypothetical bias in choice experiments, *American Journal of Agricultural Economics* 95 (5), 1136-1154.

Ehmke, M. D., J. L. Lusk, and J. A. List (2008), Is hypothetical bias a universal phenomenon? A multinational investigation, *Land Economics* 84 (3), 489-500.

European Commission (2021), Commission puts forward new strategy to make the EU's financial system more sustainable and proposes new European Green Bond Standard, https://ec.europa.eu/commission/presscorner/detail/en/ip_21_3405 (last accessed January 12, 2022).

Eurostat (2022), International Standard Classification of Education (ISCED), https://ec.europa.eu/eurostat/statistics-explained/index.php?title=International_Standard_Classification_of_Education_(ISCED) (last accessed January 12, 2022).

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018), Global evidence on economic preferences, *The Quarterly Journal of Economics* 133 (4), 1645-1692

Farrow, K., G. Grolleau, and L. Ibanez (2017), Social norms and pro-environmental behavior: A review of the evidence, *Ecological Economics* 140, 1-13.

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

Gutsche, G. and A. Ziegler (2019), Which private investors are willing to pay for sustainable investments? Empirical evidence from stated choice experiments, *Journal of Banking and Finance* 102, 1155-1182.

Haghani, M., M. Bliemer, J. Rose, H. Oppewal, and E. Lancsar (2021), Hypothetical bias in stated choice experiments: Part I. Macro-scale analysis of literature and integrative synthesis of empirical evidence from applied economics, experimental psychology and neuroimaging, *Journal of Choice Modelling* 41, 100309.

Hartzmark, S. M. and A. B. Sussman (2019), Do investors value sustainability? A natural experiment examining ranking and fund flows, *The Journal of Finance* 74 (6), 2789-2837.

Johansson-Stenman, O., and H. Svedsäter (2012), Self-image and valuation of moral goods: Stated versus actual willingness to pay, *Journal of Economic Behavior & Organization* 84 (3), 879-891.

Hess, S. and D. Palma (2019), Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application, *Journal of Choice Modelling* 32, 100170.

Hoyos, D. (2010), The state of the art of environmental valuation with discrete choice experiments, *Ecological Economics* 69 (8), 1595–1603.

Johnston, R. J., K. J. Boyle, W. Adamowicz, J. Bennett, R. Brouwer, T. A. Cameron, ... and C. A. Vossler (2017), Contemporary guidance for stated preference studies, *Journal of the Association of Environmental and Resource Economists*, 4 (2), 319-405.

Kirchler, M., F. Lindner, and U. Weitzel (2018), Ranking and risk-taking in the finance industry, *The Journal of Finance* 73 (5), 2271-2302.

Lagerkvist, C. J., A. K. Edenbrandt, I. Tibbelin, and Y. Wahlstedt (2020), Preferences for sustainable and responsible equity funds-A choice experiment with Swedish private investors, *Journal of Behavioral and Experimental Finance* 28.

List, J. A., and C. A. Gallet (2001), What experimental protocol influence disparities between actual and hypothetical stated values? Evidence from a meta-analysis, *Environmental and Resource Economics* 20 (3), 241–254.

Loomis, J. B. (2014), 2013 WAEA keynote address: Strategies for overcoming hypothetical bias in stated preference surveys, *Journal of Agricultural and Resource Economics* 34-46.

Lusardi, A. and O. S. Mitchell (2008), Planning and financial literacy: How do women fare? *American Economic Review: Papers & Proceedings* 98 (2), 413-417.

Lusk, J. L. and F. B. Norwood (2009), Bridging the gap between laboratory experiments and naturally occurring markets: an inferred valuation method, *Journal of Environmental Economics and Management* 58 (2), 236-250.

Lusk, J. L., L. McLaughlin, and S. R. Jaeger (2007), Strategy and response to purchase intention questions, *Marketing letters* 18 (1), 31-44.

Lusk, J. L. and T. C. Schroeder (2004), Are choice experiments incentive compatible? A test with quality differentiated beef steaks, *American Journal of Agricultural Economics* 86 (2), 467-482.

McFadden, D. (1973), Conditional logit analysis of qualitative choice models, In: *Frontiers in Econometrics*. Academic Press, New York, 105-142.

McFadden, D., and K. Train (2000), Mixed MNL models for discrete response, *Journal of Applied Econometrics* 15 (5), 447-470.

Menapace, L. and R. Raffaelli (2020), Unraveling hypothetical bias in discrete choice experiments, *Journal of Economic Behavior & Organization* 176, 416-430.

Mitani, Y., and N. E. Flores (2014), Hypothetical bias reconsidered: Payment and provision uncertainties in a threshold provision mechanism, *Environmental and Resource Economics*, 59 (3), 433-454.

Mørkbak, M. R., S. B. Olsen, and D. Campbell (2014), Behavioral implications of providing real incentives in stated choice experiments, *Journal of Economic Psychology*, 45, 102-116.

Morningstar (2022), Morningstar Sustainability Rating Methodology, https://www.morningstar.com/content/dam/marketing/shared/research/methodology/744156_Morningstar_Sustainability_Rating_for_Funds_Methodology.pdf (last accessed January 25th, 2022)

Moser, R., R. Raffaelli, and S. Notaro (2014), Testing hypothetical bias with a real choice experiment using respondents' own money, *European Review of Agricultural Economics* 41 (1), 25-46.

Murphy, J. J., P. G. Allen, T. H. Stevens, and D. Weatherhead (2005), A meta-analysis of hypothetical bias in stated preference valuation, *Environmental and Resource Economics* 30 (3), 313–325.

Norwood, F. B., and J. L. Lusk (2011), Social desirability bias in real, hypothetical, and inferred valuation experiments, *American Journal of Agricultural Economics* 93 (2), 528-534**.**

Paulhus, D. K. (1984), Two-component models of socially desirable responding, *Journal of Personality and Social Psychology* 46 (3), 598-609.

Paulhus, D. L. (1991), Measurement and control of response bias, In: Measures of personality and social psychological attitudes (Eds: J. P. Robinson, P. R. Shaver, and L. S. Wrightman), 17–59. New York: Academic Press.

Penn, J. M., W. Hu, and L. J. Cox (2019), The effect of forced choice with constant choice experiment complexity, *Journal of Agricultural and Resource Economics*, 44 (1835-2019-1570), 439-455.

Revelt, D. and K. Train (1998), Mixed logit with repeated choices: Households' choices of appliance efficiency level, *Review of Economics and Statistics* 80 (4), 647-657.

Riedl, A., and P. Smeets (2017), Why do investors hold socially responsible mutual funds?, *The Journal of Finance*, 72 (6), 2505-2550.

Rossi, M., D. Sansone, A. van Soest, and C. Torricelli (2019), Household preferences for socially responsible investments, *Journal of Banking and Finance* 105, 107-120.

Sanjuán-López, A. I., and H. Resano-Ezcaray (2020), Labels for a local food speciality product: The case of saffron, *Journal of Agricultural Economics* 71 (3), 778-797.

Schmidt, J., and T. H. Bijmolt (2020), Accurately measuring willingness to pay for consumer goods: a meta-analysis of the hypothetical bias, *Journal of the Academy of Marketing Science* 48 (3), 499-518.

Train, K. and M. Weeks (2005), Discrete choice models in preference space and willingness-to-pay space, In: *Applications of simulation methods in environmental and resource economics*. Springer, Dordrecht, 1-16.

Train, K. (2009), Discrete choice methods with simulation. Cambridge University Press, Cambridge, New York.

Zawojska, E., and M. Czajkowski (2017), Re-examining empirical evidence on stated preferences: importance of incentive compatibility, *Journal of Environmental Economics and Policy* 6 (4), 374-403.

Ziegler, A. (2017), Political orientation, environmental values, and climate change beliefs and attitudes: An empirical cross country analysis, *Energy Economics* 63, 144-153.

Ziegler, A. (2019), The relevance of attitudinal factors for the acceptance of energy policy measures: A micro-econometric analysis, *Ecological Economics* 157, 129-140.

Tables

Table 1: Attributes in the choice experiment

Attribute	Description	Range
Strength of sustainability	Individual bond funds differ in the extent to which sustainability criteria (i.e. environmental, social and/or governance criteria) are included in addition to financial criteria in their composition or construction. The strength of sustainability varies between "very low" and "very high" on a five-point scale based on an assessment carried out by a company.	Very low, rather low, rather high, very high
Annual returns in the past two years	The bond funds differ in terms of their average return in the years 2019 and 2020 in %.	-0.02% - 12.75%
Share of issuers of bonds from the European Union	: The bond funds differ in the countries from which the issuers of the invested bonds (e.g., corporate or public) originate. The percentage share of countries of the European Union (EU) is indicated. The remaining percentage relates to issuers of bonds not from the EU, as well as cash and other financial products.	0.00% - 83.23%
Fees:	The bond funds differ in the amount of fees (in % of the investment amount) that can be incurred during the one-year investment period.	0.00% - 6.67%

Table 2: Sample characteristics

	Germany					France		
	Full sample	Incentivized choice and no safe option group ($T1_G$)	Hypothetical choice and no safe option group ($T2_G$)	Incentivized choice and safe option group ($T3_G$)	Hypothetical choice and safe option ($T4_G$)	Full sample	Incentivized choice and no safe option group ($T1_F$)	Hypothetical choice and no safe option group ($T2_F$)
Individual characteristics								
Self-deceptive enhancement	0.75	0.74	0.77	0.75	0.72	0.79	0.80	0.77
Impression management	0.61	0.57	0.64	0.66	0.57	0.74	0.74	0.74
Financial literacy	0.56	0.58	0.52	0.56	0.57	0.66	0.68	0.63
Knows sustainable investments	0.58	0.61	0.56	0.58	0.57	0.45	0.48	0.43
High perceived risk	0.27	0.27	0.24	0.27	0.30	0.28	0.29	0.27
Warm glow	0.80	0.80	0.81	0.78	0.78	0.84	0.84	0.84
Expectation social environment	0.38	0.40	0.36	0.38	0.38	0.55	0.55	0.55
Volunteering	0.33	0.34	0.35	0.32	0.30	0.31	0.32	0.30
Social policy identification	0.63	0.67	0.60	0.63	0.63	0.47	0.45	0.48
Ecological policy identification	0.49	0.48	0.51	0.49	0.49	0.56	0.55	0.56
Female	0.36	0.35	0.37	0.34	0.37	0.42	0.44	0.40
Age	48.93	48.63	49.38	49.87	47.76	46.76	46.34	47.16
High education	0.35	0.37	0.31	0.34	0.36	0.32	0.34	0.30
Married	0.63	0.63	0.65	0.63	0.61	0.68	0.70	0.66
Western Germany	0.77	0.77	0.78	0.80	0.75	-	-	-
Experiment-related variables								
Choice certainty	0.62	0.64	0.59	0.63	0.65	0.45	0.48	0.43
Strategic behavior	0.30	0.30	0.27	0.30	0.31	0.17	0.16	0.18
Perceived winning probability	0.13	0.28	n.a.	0.20	n.a.	0.15	0.32	n.a.
Duration*	247.5	265	234	255	235	245	259	227.5
Read sustainability description again	0.06	0.07	0.05	0.05	0.05	0.22	0.26	0.18
Impact	0.47	0.48	0.48	0.47	0.43	0.36	0.40	0.32
Number of respondents	1324	408	408	253	255	829	407	422

Note: This table reports the means for all variables except for the duration.

* The median is reported for the duration, as the mean time is skewed by outliers that needed exceptionally long for the experiment (for example, due to a break in answering the survey during the experiment).

Table 3: Estimation results for treatment effects on strength of sustainability across all respondents

	Germany				France	
	No safe option groups ($T1_G$+$T2_G$)		Safe option groups ($T3_G$+$T4_G$)		No safe option groups ($T1_F$+$T2_F$)	
	Estimated mean of WTP (z-statistic)	Estimated standard deviation of WTP (z-statistic)	Estimated mean of WTP (z-statistic)	Estimated standard deviation of WTP (z-statistic)	Estimated mean of WTP (z-statistic)	Estimated standard deviation of WTP (z-statistic)
Strength of sustainability x $T2_G$ (hypothetical choice)	-0.45** (-3.44)			--	--	--
Strength of sustainability x $T4_G$ (hypothetical choice)	--	--	-0.14 (-1.01)		--	--
Strength of sustainability x $T2_F$ (hypothetical choice)	--	--	--	--	-0.26 (-1.54)	
Strength of sustainability	2.35*** (15.46)	1.08*** (10.18)	1.95*** (12.21)	1.51*** (13.78)	1.96*** (12.24)	1.54*** (17.10)
Annual returns in the past two years	0.52*** (11.81)	0.48*** (13.92)	0.45*** (10.05)	0.50*** (11.52)	0.68*** (13.69)	0.45*** (15.85)
Share of bond issuers from the EU	0.02*** (5.01)	0.00*** (13.07)	0.01*** (2.56)	0.04*** (6.98)	0.02*** (4.56)	0.08*** (14.23)
Fees	-1.01*** (-14.51)	1.34*** (13.71)	-0.70*** (-9.51)	0.83*** (7.50)	-1.52*** (-20.85)	1.33*** (12.72)
Safe option	--	--	0.27 (0.31)	6.80*** (10.83)	--	--
Number of respondents (number of choices)	816 (4,896)		508 (3,048)		829 (4,974)	

Note: Simulated maximum likelihood estimates (z-statistics) in mixed logit models in willingness to pay space with correlated coefficients. The first model is estimated for the N = 816 respondents (and thus 816 × 6 = 4,896 observations) with 408 respondents from the incentivized choice and no safe option group and 408 respondents from the hypothetical choice and no safe option treatment in Germany. The second model is estimated for the N = 508 respondents (and thus 508 × 6 = 3,084 observations) with 253 respondents from the incentivized choice and safe option group and 255 respondents from the hypothetical choice and safe option group in Germany. The third model is estimated for the N = 829 (and thus 816 × 6 = 4,974 observations) with 407 respondents from the incentivized choice and no safe option group and 422 respondents from the hypothetical choice and no safe option group in France. The "estimated mean of WTP" columns report the estimated mean willingness to pay estimates of the respective non-price explanatory variables and the estimated mean of the log of the fees coefficients. The "estimated standard deviation of WTP" columns report the estimated standard deviations of the willingness to pay distribution of the random parameters. The dependent variable is choice in all considered models. As explanatory variables, we consider the random parameters for strength of sustainability, annual returns in the past two years, share of bond issuers from the EU, and safe option in all models. The interactions between the strength of sustainability and the treatment group variables are assumed to be fixed. In the second model, we additionally consider a random parameter for the safe option. All variables are defined in Section 2.2. * (**, ***) means that the respective effect is different from zero at the 10% (5%, 1%) significance level, respectively.

Table 4a: Estimation results for treatment effects on strength of sustainability across different subsamples in the no safe option groups in Germany ($T1_G$+$T2_G$)

	Self-deceptive enhancement		Impression management		Strategic behavior		Choice certainty	
	High	Low	High	Low	High	Low	High	Low
Estimated mean of WTP (z-statistic)								
Strength of sustainability × $T2_G$ (hypothetical choice)	-0.51 (-0.98)	-0.18 (-1.59)	-0.49*** (-3.00)	-0.44** (-2.17)	-0.16 (-0.57)	-0.58*** (-8.69)	-0.52*** (-3.61)	-0.09 (-0.36)
Strength of sustainability	2.44*** (8.80)	1.95*** (10.17)	2.54*** (12.84)	2.07*** (11.56)	2.07*** (8.91)	2.75*** (22.46)	2.43*** (16.67)	2.19*** (7.88)
Annual return in the past two years	0.52*** (10.03)	0.39*** (13.98)	0.46*** (11.37)	0.48*** (7.14)	0.51*** (8.46)	0.54*** (16.99)	0.56*** (15.35)	0.53*** (7.49)
Share of bond issuers from the EU	0.02*** (4.47)	0.00 (1.00)	0.02*** (3.62)	0.03*** (3.94)	0.01 (1.44)	0.04*** (9.09)	0.03*** (6.92)	-0.01 (-1.34)
Fees	-1.03*** (-12.48)	-0.92*** (-5.68)	-0.99*** (-11.68)	-1.01*** (-9.10)	-0.71*** (-8.32)	-1.31*** (-14.18)	-0.66*** (-6.95)	-1.62*** (-12.58)
Estimated standard deviation of WTP (z-statistic)								
Strength of sustainability	1.04*** (4.11)	0.52*** (6.83)	0.96*** (8.34)	1.24*** (7.55)	0.85*** (2.71)	0.83*** (17.75)	0.92*** (9.09)	1.58*** (9.66)
Annual return in the past two years	0.47*** (8.64)	0.45*** (28.17)	0.41*** (11.83)	0.59*** (11.88)	0.44*** (5.93)	0.48*** (24.42)	0.42*** (16.67)	0.05 (1.46)
Share of bond issuers from the EU	0.00*** (10.11)	0.06*** (11.85)	0.00*** (8.19)	0.07*** (8.92)	0.00*** (6.29)	0.05*** (16.63)	0.06*** (22.29)	0.00 (0.03)
Fees	1.14*** (10.40)	0.91** (2.02	0.83*** (3.78)	1.33*** (6.62)	0.66*** (4.69)	1.20*** (6.35)	1.26*** (5.83)	1.16*** (4.51)
Number of observations (number of choices)	614 (3,684)	202 (1,212)	495 (2,970)	321 (1,926)	237 (1,422)	579 (3,474)	655 (2,991)	557 (1,905)

Note: Simulated maximum likelihood estimates (z-statistics) in mixed logit models in willingness to pay space with correlated coefficients. The first and second models are estimated for the N = 614 respondents (and thus 614 × 6 = 3,684 observations) from the incentivized choice and no safe option group and from the hypothetical choice and no safe option treatment in Germany with high self-deceptive enhancement and for the N = 202 respondents (and thus 202 × 6 = 1,212 observations) from the incentivized choice and no safe option group and from the hypothetical choice and no safe option treatment in Germany with low self-deceptive enhancement. The third and fourth models are estimated for the N = 495 respondents (and thus 495 × 6 = 2,970 observations) from the incentivized choice and no safe option group and from the hypothetical choice and no safe option treatment in Germany with high impression management and for the N = 321 respondents (and thus 321 × 6 = 1,926 observations) from the incentivized choice and no safe option group and from the hypothetical choice and no safe option treatment in Germany with low impression management. The fifth and sixth models are estimated for the N = 237 respondents (and thus 237

× 6 = 1,422 observations) from the incentivized choice and no safe option group and from the hypothetical choice and no safe option treatment in Germany who mention the analysis of preferences for sustainable investments as the goal of this study and for the N = 579 respondents (and thus 579 × 6 = 3,474 observations) from the incentivized choice and no safe option group and from the hypothetical choice and no safe option treatment in Germany who did not mention the analysis of preferences for sustainable investments as the goal of this study. The seventh and eighth models are estimated for the N = 655 respondents (and 2,991 observations) from the incentivized choice and no safe option group and from the hypothetical choice and no safe option treatment in Germany where only choices are included in which respondents stated to have been rather or very certain about their choice and for the N = 557 respondents (and 1,905 observations) from the incentivized choice and no safe option group and from the hypothetical choice and no safe option treatment in Germany where only choices are included in which respondents stated not to have been rather or very certain about their choice. The "estimated mean of WTP" columns report the estimated mean willingness to pay estimates of the respective non-price explanatory variables and the estimated mean of the log of the fees coefficients. The "estimated standard deviation of WTP" columns report the estimated standard deviations of the willingness to pay distribution of the random parameters. The dependent variable is choice in all considered models. As explanatory variables, we consider the random parameters for strength of sustainability, annual returns in the past two years, share of bond issuers from the EU, and safe option in all models. The interactions between the strength of sustainability and the treatment group variables are assumed to be fixed. In the second model, we additionally consider a random parameter for the safe option. All variables are defined in Sections 2.2 and 2.3. * (**, ***) means that the respective effect is different from zero at the 10% (5%, 1%) significance level, respectively.

Table 4b: Estimation results for treatment effects on strength of sustainability across different subsamples in the no safe option groups in Germany ($T1_G$+$T2_G$)

	Financial literacy		Knows sustainable investments		Impact		Winning probability	
	High	Low	High	Low	High	Low	High	Low
Estimated mean of WTP (z-statistic)								
Strength of sustainability ×	-0.42**	-0.12	-0.30	-0.56**	-0.40	-0.42***	-0.71	-0.36***
$T2_G$ (hypothetical choice)	(-2.52)	(-0.22)	(-1.41)	(-2.21)	(-1.42)	(-2.99)	(-1.18)	(-3.29)
Strength of sustainability	2.19***	2.64***	2.39***	1.96***	3.90***	1.47***	2.83***	2.26***
	(13.46)	(9.43)	(11.66)	(8.84)	(13.71)	(9.33)	(5.61)	(13.79)
Annual return in the past two years	0.60***	0.28	0.53***	0.40***	0.51***	0.55***	0.55	0.49***
	(12.57)	(1.58)	(11.35)	(4.96)	(11.10)	(10.08)	(0.85)	(13.88)
Share of bond issuers from the EU	0.03***	0.01	0.03***	0.02***	0.06***	0.01	0.03***	0.02***
	(6.01)	(0.72)	(5.47)	(2.63)	(11.35)	(1.61)	(3.87)	(4.75)
Fees	-0.59***	-1.69***	-0.77***	-1.28***	-1.30***	-0.89***	-1.22***	-0.93***
	(-7.60)	(-13.66)	(-8.31)	(-12.81)	(-12.09)	(-9.84)	(-3.13)	(-12.18)
Estimated standard deviation of WTP (z-statistic)								
Strength of sustainability	0.98***	1.16***	1.01***	0.72***	1.26***	0.63***	1.17	0.98***
	(9.37)	(5.06)	(7.86)	(4.18)	(18.76)	(6.37)	(1.03)	(10.18)
Annual return in the past two years	0.36***	0.70***	0.44***	0.53***	0.46***	0.49***	0.52***	0.44***
	(13.17)	(5.47)	(15.33)	(9.13)	(16.29)	(13.39)	(5.09)	(15.93)
Share of bond issuers from the EU	0.00***	0.06***	0.05***	0.05***	0.06***	0.04***	0.04***	0.01***
	(12.49)	(11.29)	(10.35)	(6.30)	(14.68)	(9.01)	(6.75)	(13.16)
Fees	0.91***	1.22***	1.15***	1.15***	0.63**	1.27***	1.35**	0.90**
	(9.36)	(3.32)	(9.07)	(6.16)	(2.02)	(6.61)	(4.13)	(2.51)
Number of observations (number of choices)	448 (2,688)	368 (2,208)	369 (2,214)	341 (2,046)	390 (2,340)	426 (2,556)	523 (3,138)	701 (4,206)

Note: Simulated maximum likelihood estimates (z-statistics) in mixed logit models in willingness to pay space with correlated coefficients. The first and second models are estimated for the N = 448 respondents (and thus 448 × 6 = 2,688 observations) from the incentivized choice and no safe option group and from the hypothetical choice and no safe option treatment in Germany with the same or higher financial literacy than the sample mean and for the N = 368 respondents (and thus 368 × 6 = 2,208 observations) from the incentivized choice and no safe option group and from the hypothetical choice and no safe option treatment in Germany with a lower financial literacy than the sample mean. The third and fourth models are estimated for the N = 369 respondents (and thus 369 × 6 = 2,214 observations) from the incentivized choice and no safe option group and from the hypothetical choice and no safe option treatment in Germany with high knowledge about sustainable investments and for the N = 341 respondents (and thus 341 × 6 = 2,046 observations) from the

incentivized choice and no safe option group and from the hypothetical choice and no safe option treatment in Germany with low or mediocre knowledge about sustainable investments. The fifth and sixth models are estimated for the N = 390 respondents (and thus 390 × 6 = 2,340 observations) from the incentivized choice and no safe option group and from the hypothetical choice and no safe option treatment in Germany who rather or strongly agree that bond funds can contribute to a sustainable development and for the N = 426 respondents (and thus 426 × 6 = 2,556 observations) from the incentivized choice and no safe option group and from the hypothetical choice and no safe option treatment in Germany who do not rather or strongly agree that bond funds can contribute to a sustainable development. The seventh and eighth models are estimated for the N = 523 respondents (and thus 523 × 6 = 3,138 observations) from the incentivized choice and no safe option group and from the hypothetical choice and no safe option treatment in Germany where only those respondents from the incentivized choice and no safe option group are included who state that their winning probability is rather or very high and for the N = 701 respondents (and thus 701 × 6 = 4,206 observations) from the incentivized choice and no safe option group and from the hypothetical choice and no safe option treatment in Germany where only those respondents from the incentivized choice and no safe option group are included who do not state that their winning probability is rather or very high. The "estimated mean of WTP" columns report the estimated mean willingness to pay estimates of the respective non-price explanatory variables and the estimated mean of the log of the fees coefficients. The "estimated standard deviation of WTP" columns report the estimated standard deviations of the willingness to pay distribution of the random parameters. The dependent variable is choice in all considered models. As explanatory variables, we consider the random parameters for strength of sustainability, annual returns in the past two years, share of bond issuers from the EU, and safe option in all models. The interactions between the strength of sustainability and the treatment group variables are assumed to be fixed. In the second model, we additionally consider a random parameter for the safe option. All variables are defined in Sections 2.2 and 2.3. * (**, ***) means that the respective effect is different from zero at the 10% (5%, 1%) significance level, respectively.

Table 5: Estimation results for effects of individual characteristics on strength of sustainability in the no safe option groups in Germany ($T1_G$+$T2_G$)

	Germany			
	Incentivized choice and no safe option group ($T1_G$)		Hypothetical choice and no safe option ($T2_G$)	
	Estimated mean of WTP (z-statistic)	Estimated standard deviation of WTP (z-statistic)	Estimated mean of WTP (z-statistic)	Estimated standard deviation of WTP (z-statistic)
Strength of sustainability x high perceived risk	-0.33*		-0.35*	
	(-1.74)		(-1.76)	
Strength of sustainability x warm glow	0.90***		0.94***	
	(4.24)		(4.43)	
Strength of sustainability x expectation social environment	-0.01		0.17	
	(-0.05)		(0.84)	
Strength of sustainability x volunteering	-0.18		0.20	
	(-1.25)		(1.09)	
Strength of sustainability x high social policy orientation	0.04		0.06	
	(0.22)		(0.30)	
Strength of sustainability x high ecological policy orientation	0.78***		0.96***	
	(4.82)		(4.34)	
Strength of sustainability x female	0.06		-0.01	
	(0.37)		(-0.06)	
Strength of sustainability x age	-0.00		-0.00	
	(-0.89)		(-0.60)	
Strength of sustainability x high education	0.26		-0.32*	
	(1.25)		(-1.86)	
Strength of sustainability x living together or married	0.06		0.36*	
	(0.37)		(1.88)	
Strength of sustainability x Western Germany	0.21		0.21**	
	(1.14)		(2.01)	
Strength of sustainability	0.93**	0.69***	0.70***	0.96***
	(2.42)	(11.90)	(1.81)	(7.88)
Annual returns in the past two years	0.54***	0.46***	0.56***	0.46***
	(9.44)	(14.82)	(10.41)	(10.14)
Share of bond issuers from the EU	0.03***	0.04***	0.01***	0.06***
	(4.70)	(10.70)	(3.77)	(9.63)
Fees	-1.11***	1.25***	-0.93***	1.28***
	(-11.05)	(9.13)	(-9.84)	(9.26)
Number of respondents (number of choices)	408 (2,448)		408 (2,448)	

Note: Simulated maximum likelihood estimates (z-statistics) in mixed logit models in willingness to pay space with correlated coefficients. The first model is estimated for the N = 408 respondents (and thus 408 × 6 = 2,448 observations) from the incentivized choice and no safe option group in Germany. The second model is estimated for the N = 408 respondents (and thus 408 × 6 = 2,448 observations) from the hypothetical choice and no safe option treatment group in Germany. The "estimated mean of WTP" columns report the estimated mean willingness to pay estimates of the respective non-price explanatory variables and the estimated mean of the log of the fees coefficients. The "estimated standard deviation of WTP" columns report the estimated standard deviations of the willingness to pay distribution of the random parameters. The dependent variable is choice in all considered models. As explanatory variables, we consider the random parameters for strength of sustainability, annual returns in the past two years, share of bond issuers from the EU, and safe option in all models. The interactions between the strength of sustainability and the individual characteristics are assumed to be fixed. All variables are defined in Sections 2.2 and 2.3. * (**, ***) means that the respective effect is different from zero at the 10% (5%, 1%) significance level, respectively.

Appendix A: Additional tables

Table A.1: Randomization checks

	Difference in means (z-statistics)						
	$T1_G$ versus $T2_G$	$T1_G$ versus $T3_G$	$T1_G$ versus $T4_G$	$T2_G$ versus $T3_G$	$T2_G$ versus $T4_G$	$T3_G$ versus $T4_G$	$T1_F$ versus $T2_F$
Self-deceptive enhancement	0.97	0.26	-0.46	-0.59	-1.32	-0.65	-1.34
Impression management	1.94*	2.12**	-0.02	0.43	-1.73*	-1.94*	0.07
Financial literacy	-1.83*	-0.50	-0.31	1.10	1.29	0.17	-1.67*
Knows sustainable investments	-1.49	-0.78	-1.00	0.52	0.31	-0.19	-1.45
High perceived risk	-1.12	-0.05	0.65	0.93	1.65*	0.63	-0.56
Warm glow	0.18	-0.70	-0.53	-0.86	-0.69	0.15	0.04
Expectation social environment	-1.08	-0.55	-0.43	0.39	0.52	0.11	0.12
Volunteering	0.07	-0.61	-1.10	-0.67	-1.16	-0.44	-0.65
Social policy identification	-2.11**	-1.00	-0.93	0.84	0.92	0.07	0.90
Ecological policy identification	0.70	0.08	0.28	-0.53	-0.33	0.18	0.19
Female	0.61	0.89	-0.62	0.36	-1.16	-1.35	0.77
Age	0.58	-0.28	0.58	-0.79	0.06	0.77	-1.15
High education	-1.55	-0.56	-0.01	0.80	1.35	0.49	-1.40
Married	0.66	-0.10	-0.63	-0.68	-1.21	-0.48	-1.20
Western Germany	0.42	0.87	-0.60	0.51	-0.98	-1.33	-
Number of respondents	816	663	661	663	661	508	827

Note: * (**, ***) means that the difference in the means between the treatment groups on the basis of a mean comparison z-test is different from zero at the 10 % (5 %, 1 %) significance level, respectively.

Table A.2: Treatment effects on choice certainty

Dependent variable:	Choice certainty	
Country:	Germany	France
Parameter:	Mean	Mean
Treatment (base: incentivized and no safe option)		
Hypothetical and no safe option	-0.22* (-1.91)	-0.19* (-1.74)
Incentivized and safe option	-0.02 (-0.20)	
Hypothetical and safe option	0.04 (0.29)	
Respondents	1,324	829

Notes: Maximum likelihood estimates (z-statistics) in the binary logit model when no safe option is provided with choice certainty as dependent variable that takes the value one when a respondent was rather or very certain about his or her choice in the respective choice set. * (**, ***) means that the respective effect is different from zero at the 10% (5%, 1%) significance level, respectively. Standard errors clustered for each individual as each individual answered the choice certainty question six times.