The Review of Financial Studies

Do Investors Care about Impact?

Florian Heeb
MIT Sloan, USA and University of Zürich, Switzerland

Julian F. Kölbel
University of St. Gallen, Switzerland and MIT Sloan, USA

Falko Paetzold
EBS University, Germany and University of Zürich, Switzerland

Stefan Zeisberger
Radboud University, the Netherlands and University of Zürich, Switzerland

We assess how investors' willingness-to-pay (WTP) for sustainable investments responds to the social impact of those investments, using a framed field experiment. While investors have a substantial WTP for sustainable investments, they do not pay significantly more for more impact. This also holds for dedicated impact investors. When investors compare several sustainable investments, their WTP responds to relative, but not to absolute, levels of impact. Regardless of investments' impact, investors experience positive emotions when choosing sustainable investments. Our findings suggest that the WTP for sustainable investments is primarily driven by an emotional, rather than a calculative, valuation of impact. (*JEL* D62, G11, G41, Q56)

Received January 17, 2022; editorial decision July 12, 2022 by Editor Tarun Ramadorai.

We thank Tarun Ramadorai and two anonymous reviewers for their thoughtful comments. Further, we thank Erol Bilecen, Timo Busch, Marc Chesney, Alain Cohn, Kornelia Fabisik, Jonathan Harris, Sam Hartzmark, Thorsten Hens, Coralie Jaunin, Pascal Kieren, Christoph Merkle, Jan Schmitz, Zacharias Sautner, Julien Senn, Hersh Shefrin, Paul Smeets, Janneke Toussaint, Michael Ungeheuer, Alexander Wagner, Utz Weitzel, Ziwei Zhao, and participants in seminars and conferences at the AFA Annual Meeting, ARCS, the Dutch Ministry of Finance, the GRASFI, the JEF Seminar, the Liechtenstein Sustainable Finance Workshop, the MPI Bonn, Virtual Experimental Finance Workshop, the SEF Annual Conference, the DGF, Bonn Household Finance Workshop, Radboud University, the SFI Research Days, the SGF, the University of Exeter, and the University of Zurich for helpful comments and suggestions. We gratefully acknowledge financial and participant pool support from the VEB. Florian Heeb was supported by a PRI FIR research award grant. We further thank the VEB and Armand Kersten, Toniic and Kristin Siegel, NEXUS and Tharald Nustad, Credit Suisse and James Gifford, Bank Vontobel and Daniel Grüneisen, Bank Julius Baer and Alessandro Anastassio, Tiedemann Advisors and Brad Harrison, and VALUEworks and Peter Wüthrich for providing support regarding data collection. Send correspondence to Julian F. Kölbel, University of St. Gallen, Unterer Graben 21, 9000 St.Gallen, Switzerland; Telephone: +41 71 224 7161; E-mail: julian.koelbel@unisg.ch.

The Review of Financial Studies 36 (2023) 1737–1787

https://doi.org/10.1093/rfs/hhac066 Advance Access publication September 13, 2022

Sustainable investing is seen as a mechanism for curbing negative externalities (Pastor, Stambaugh, and Taylor 2021; Oehmke and Opp 2019; Broccardo, Hart, and Zingales 2020; Benabou and Tirole 2010; Landier and Lovo 2020; Hong, Wang, and Yang 2021). This mechanism is based on the assumption that a substantial pool of investors hold prosocial preferences. Indeed, recent research has demonstrated that prosocial preferences affect investment decisions (Riedl and Smeets 2017), and that these preferences result in sizable, market-wide fund flows toward sustainable investments (Hartzmark and Sussman 2019). Some investors show an explicit willingness-to-pay (WTP) for investments with social "impact" (Barber, Morse, and Yasuda 2021), which we define as a positive externality of an investment.

Standard decision theory would predict that investors are consequentialists—so that the utility that investors derive from a sustainable investment is proportional to the impact of that investment.[1] This consequentialist view is also adopted widely in current models of sustainable investing. Yet research on charitable giving and public good valuation shows that individuals often display scope insensitivity—that is, they are insensitive to the magnitude of their impact.[2] Scope insensitivity may be prominent in sustainable investing decisions. If it is, this could undermine the effectiveness of sustainable finance as a whole, as the financial industry may not have an incentive to supply products with substantial impact. To shed light on this issue, we ask the question: How does investors' WTP for sustainable investments respond to the impact of these investments?

We investigate this research question in a preregistered, framed field experiment[3] with experienced investors. They choose between a sustainable investment with a quantified impact and a financially equivalent investment with zero impact. The investment choices are incentivized and consequential—that is, investors can make real money, and their choices have real impact. Based on the investors' choices, we elicit their WTP for the sustainable investment in terms of additional fees they are willing to pay. We operationalize impact in the form of carbon dioxide (CO_2) emissions reductions and vary the impact of the sustainable investment by a factor of 10 between two treatment groups. We run the experiment in different variations with a panel of 527 experienced private investors as well as with a unique panel of 125 dedicated high-net-worth impact investors whom we recruited via impact investor networks and specialized wealth managers.

Our main experiment provides evidence that investors' WTP does not respond significantly to the level of impact that a sustainable investment offers. As a starting point, we confirm that investors are willing to pay a substantial

[1] Traditional models of altruistic behavior often assume that individuals contribute to public goods because they derive utility from the well-being of others (Becker 1974; Eckel and Grossman 1996; Andreoni and Miller 2002).

[2] See, for example, Karlan and Wood (2017), Metzger and Günther (2019), Desvousges et al. (1992), and Kahneman and Knetsch (1992).

[3] According to the classification of Harrison and List (2004).

amount for a sustainable investment with some impact. However, investors' WTP does not significantly differ between an investment that saves 0.5 tons of CO_2 emissions and one that saves 5 tons. The experiment does not rule out that some investors are sensitive to impact, but in aggregate that seems to be a second-order effect. In sum, we find that, although investors care whether an investment has an impact or not, they hardly care about the magnitude of that impact.

We take several measures to ensure the robustness of this finding. First, we make sure that investors intuitively understand what a ton of CO_2 means and that the investment's impact is salient when investors make their choices. Most investors (95%) can exactly recall the impact of the investment after the choice experiment. Second, we run a series of variations of our main experiment. Using the same subject pool, we vary the investments' past financial performance instead of their impact, assuming that investors care about the level of financial performance. We can demonstrate that in this setting, investors' WTP is highly sensitive to differences in financial performance. Using an additional sample of 2,800 participants recruited via Amazon Mechanical Turk (MTurk), we first replicate our findings with a larger number of participants (n=1,000). Subsequently, we rely on MTurk participants to run eight variations of the main experiment to probe for the potential influence of experimenter demand and our elicitation method. The result of no significant difference between impact treatments holds across all MTurk experiments. Finally, relying on a sample of 554 university students, we show that our results hold both before and after the COVID-19 crisis.

Extending the main finding, we explore the reasons for the observed insensitivity to impact. First, our results may be due to the fact that investors lack experience in dealing with impact. To test this hypothesis, we repeat the experiment, but this time with a unique sample of high-net-worth impact investors who have substantial experience with impact investing. We find that investors in this sample are just as insensitive to impact as are the private investors in the main experiment. This suggests that the observed insensitivity to impact is not driven by a lack of experience, and that it is unlikely to disappear as more investors gain experience with sustainable investing.

Second, we investigate whether the ability to directly compare impact information increases investors' sensitivity to impact. To this end, we run our experiment in a joint evaluation setup, in which investors receive information on the impact of each of two sustainable investments juxtaposed. Again, the impact of the investments differs by a factor of 10. Yet investors' WTP is only 28% higher for the high-impact investment in this setting, even though it has a 900% higher impact. Thus, while investors show some reaction to impact information in direct comparison, their WTP per unit of impact remains inconsistent. Analyzing sub-groups, we find that, even in the joint evaluation design, one-third of investors are entirely insensitive to the level of impact. In contrast, the 30% of investors with the highest sensitivity to impact pay, on

average, 6 times more for 10 times more impact. Furthermore, we show that in the joint evaluation setup investors' WTP does not depend on the absolute level of impact. Reducing the impact of each investment by a factor of 10 barely changes investors' WTP. This indicates that investors' WTP for sustainable investments depends strongly on the available (arbitrary) choice set and not on the absolute level of impact.

Third, we explore whether the WTP for sustainable investments is driven by the emotional experience of choosing a sustainable option rather than by a calculative appraisal of impact. Relying on a postexperiment survey, we find that investors' WTP is correlated with the level of positive emotions they experience when choosing the sustainable investment. The impact of the sustainable investment, however, does not influence these positive emotions. In a regression, we show that investors' WTP per ton of CO_2 is strongly correlated with this level of positive emotions, but not with investors' individual estimates of what it costs to save one ton of CO_2. This suggests that investors' valuation of impact is mainly driven by feelings rather than by calculation. We also explain how an emotional valuation may reconcile the results of the main experiment and of the joint evaluation extension.

Taking all our findings together, we suggest viewing the average prosocial investor as a "warm glow" optimizer,[4] rather than a consequentialist, who optimizes the impact of her investments. Our results leave open the possibility that some investors are sensitive to impact under specific circumstances, and suggest that a more calculative decision mode can be encouraged by tweaking the choice environment. However our combined results imply that a calculative appraisal of impact is second-order compared to the emotional warm glow investors derive from choosing a sustainable option. Ultimately, this suggests that the positive emotions derived from choosing sustainable investments are an important driver of the trend for sustainable investing.

Our paper contributes to the literature that investigates the influence of prosocial preferences on investment decisions (Barber, Morse, and Yasuda 2021; Riedl and Smeets 2017; Hartzmark and Sussman 2019; Bauer, Ruof, and Smeets 2021). While we confirm previous findings of a substantial WTP for sustainable investments, our results show that this WTP does not scale with the level of impact that those investments offer. Our results suggest that investors' valuation of sustainable investments is more akin to charitable giving than to financial optimization. While scope insensitivity has been shown in other contexts (Null 2011; Karlan and Wood 2017; Metzger and Günther 2019; Hsee and Rottenstreich 2004; Desvousges et al. 1992; Kahneman and Knetsch 1992), it has not been demonstrated in the context of financial decision-making, where it may well be the case that investors approach contributing to the public good in

[4] Models of warm glow or "impure" altruism focus on emotional valuation. They assume that individuals do not derive utility from the well-being of others, but from an emotional response to the act of behaving prosocially (Andreoni 1990).

a more calculating, consequentialist manner. Our results demonstrate, however, that scope insensitivity is also a relevant issue when individuals express their prosocial preference in an investment context.

Our work is also related to the literature that explores the role of emotions in financial decision-making (Finucane et al. 2000; Slovic et al. 2007; Kuhnen and Knutson 2011). Affective decision-making has been put forward as an explanation for several puzzles in financial markets, including the home bias (Huberman 2001; Coval and Moskowitz 1999; Strong and Xu 2003), or the IPOs of glamorous companies (MacGregor et al. 2000). Hartzmark and Sussman (2019) suggest that emotions may also drive investors' valuation of sustainable investments. We confirm that investors' WTP for sustainable investments is positively correlated with the positive emotions they experience when choosing a sustainable investment option. This highlights that emotions play an important role in the behavior of prosocial investors.

Finally, we challenge a key assumption in the literature that explores the effects of prosocial preferences on asset pricing. A growing number of theoretical papers model how prosocial investors influence asset prices either because they have a taste for "green" assets (Pastor, Stambaugh, and Taylor 2021; Pedersen, Fitzgibbons, and Pomorski 2021; Heinkel, Kraus, and Zechner 2001) or because they explicitly care about aggregate externalities (Broccardo, Hart, and Zingales 2020; Oehmke and Opp 2019). These models suggest that prosocial investors, by expressing their preferences in the financial market, incentivize companies to reduce externalities. In essence, these models assume that prosocial investors' utility increases with the impact of their investments, and that these investors make trade-offs between the financial performance and the impact of their investments. Our results challenge this assumption and suggest that prosocial investors are more likely to maximize financial performance while optimizing the warm glow that they derive from their choices. Modeling investor behavior in such a way would likely emphasize the importance of the structure of the sustainable investment industry, information asymmetry, and the way products are marketed to investors. Without measures in place that align the experience of warm glow with a product's underlying impact, sustainable investing may turn out to be a much less effective mechanism than previously thought for curbing externalities. Thus, future studies may yield important insights by modeling the behavior of prosocial investors in a way that explicitly reflects the role of warm glow.

Our paper is also related to three contemporary working papers. Humphrey et al. (2020) run an investment game in which investment returns are positively or negatively linked to charitable benefits. They show that investors allocate less to investment options that entail negative effects on charities, but not more to those that entail positive effects. Although their paper addresses the positive/negative dichotomy and not different levels of impact, its results are consistent with ours in the sense that investors' valuation of externalities is not

linear. Bonnefon et al. (2022) implement an auction of claims on hypothetical corporations that donate parts of their profits to charity. The authors find that respondents' WTP scales with monetary contributions to charities in a linear relation, irrespective of whether their investment is pivotal to a contribution. Brodback, Günster, and Pouget (2021) employ initial public offerings of assets that have identical financial payoffs but differ in the intensity and timing of their social responsibility aspects. The authors find that participants' WTP increases with positive externalities, implemented as donations to charity. Our findings may seem contradictory to those of Bonnefon et al. (2022) and Brodback, Günster, and Pouget (2021), as in both participants' WTP increases with the positive externalities of investment options. A key difference in the experimental setup, however, is that in both of these studies externalities are expressed in monetary units and implemented as donations. This enables participants to compare their impact one-to-one to the costs of sustainable investments. Expressing impact in monetary terms may be conducive to a calculative valuation of externalities (Hsee and Rottenstreich 2004), and it relieves investors of the difficult task of valuing impact. Our paper, meanwhile, examines the case in which externalities are not monetized, a feature that is widespread in the market for sustainable investment products. In this case, we find that WTP does not scale with the level of impact in a linear fashion, and we suggest that investors' WTP is driven by an emotional valuation. Reading the findings of these different papers together with our results suggests that monetizing impacts could be an effective measure for avoiding scope insensitivity in sustainable investing.

In terms of practical implications, our results highlight that there is a severe risk of greenwashing in sustainable finance. The market for sustainable investing is expanding quickly, in 2020 surging past a total volume of USD 35 trillion (Global Sustainable Investment Alliance 2021). This growth is raising hopes that sustainable investing might help tackle major societal challenges, such as curbing carbon emissions. However, the inconsistent WTP for impact that we demonstrate in this paper creates an incentive problem. If investors' WTP for sustainable investments scales with emotional warm glow rather than with impact, financial institutions have an incentive to create products that offer warm glow rather than impact. This is especially the case when offering impact comes at a cost.[5] Furthermore, our results suggest that financial institutions have an incentive to structure their offerings in such a way that sustainable products with little impact stand out as the most impactful option available. This may result in a market for sustainable investment products that benefits

[5] Although there are claims to the contrary, most theoretical models imply that prosocial investors need to accept lower financial performance ex ante in order to have impact (Oehmke and Opp 2019; Heinkel, Kraus, and Zechner 2001; Pastor, Stambaugh, and Taylor 2021). In addition, there are—for sustainable investments—additional requirements for data and expertise that are likely to add to management fees.

investors in terms of warm glow and financial institutions in terms of profits, but fails to fulfill its potential for solving societal problems.

1. Study Design

We address the question of investor sensitivity to impact in framed field experiments and following a preregistered experimental procedure.[6] All experiments were conducted between May and September 2020, and our robustness checks in April and May 2022. Based on a series of investment decisions, we assess investors' WTP for a sustainable investment compared to that for a "conventional" investment. Both investment options are presented as equity funds. The sustainable investment has some level of impact, while the conventional investment has zero impact but is otherwise equivalent. We vary the level of impact between subjects, allowing us to investigate how WTP responds to the level of impact. We measure investors' WTP in terms of the front-end fee that investors are willing to pay for the sustainable investment. The investment decisions are consequential, in terms of both payout and impact.

We operationalize the investments' impact in terms of carbon dioxide (CO_2) emissions savings brought about by investing in the product. We choose this measure for several reasons. First, CO_2 emissions and climate change are widely discussed in the financial press. Thus, most investors understand the topic and are familiar with the metric. Second, CO_2 emissions are clearly defined and measurable. Many metrics that are used in the context of sustainable investing are opaque in the sense that it is not clear what they measure and how they measure. In such cases, investors might not respond to impact because it is uncertain what the numbers mean. The measurement of CO_2 emissions, in contrast, is well established. Third, there are markets for CO_2 emissions savings, such as the European Emission Trading Scheme, which means that prices for CO_2 emissions are easily observable in the public domain. Finally, we are able to actually realize the impact of the sustainable investment through verified carbon emission reduction projects (see Section A.1 for details). Thus, quantifying impact in terms of CO_2 emissions should allow investors to interpret impact in a quantitative sense.

Figure 1 illustrates the flow of the experimental procedure we use to elicit investors' WTP. It proceeds in four steps: instructions, information on investments, investment decisions, and a postexperiment survey.

1.1 Instructions and incentives

In the first step, investors receive detailed instructions on the investment decisions and on investor incentivization. We make sure that participants understand how the investment decisions work and that it is in their best interest to state their true preferences (see Figure A.1).

[6] For preregistration details, see https://aspredicted.org/w5f8i.pdf.

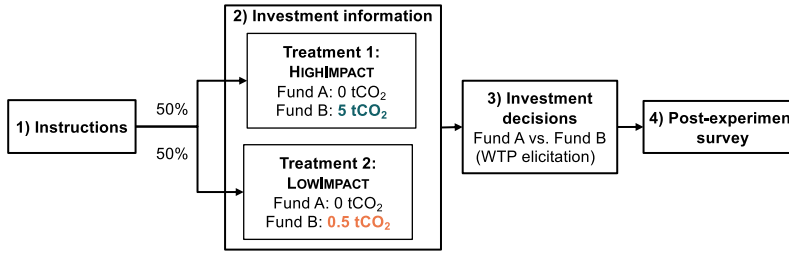


Figure 1
Experimental procedure of the main experiment
This figure provides an illustration of the four steps of the experimental procedure we use in our main experiment. In the second step, participants are randomly assigned either to the HIGHIMPACT treatment or the LOWIMPACT treatment.

The investment decisions are incentivized with relatively high stakes. For 10 randomly selected investors, we make a real €1,000 investment on their behalf, based on their investment decisions. To guarantee that participants reveal their true WTP, we apply the Becker–DeGroot–Marschak (BDM) mechanism (Becker, Degroot, and Marschak 1964), a standard procedure in the judgment and decision-making literature. Using the BDM mechanism, we determine in which option to invest and a front-end fee, which we deduct from the investment amount (see Section A.1 for the detailed procedure). After 1 year, we pay out the full value of this investment to the selected investors.

To familiarize investors with the decision procedure, we guide them through an example (see Figure A.2). We also require them to complete a brief quiz in order to check whether they understand the potential consequences of their decisions. Investors who "fail" the quiz twice receive the correct answers and a short explanation.

1.2 Investment information

In the second step, investors receive information about the financial performance and the impact of the two investment options. For each of the two investments, we provide information on the asset class, the market segment, the annualized return over the last 3 years, and the risk level according to Morningstar's risk rating, all of which are identical for both investments. For the sustainable investment option, we additionally provide information on how much CO_2 emissions a €1,000 investment saves (see Figure 2 for an example).

To make sure that respondents understand the impact information, we translate the CO_2 savings into more intuitively comprehensible figures. We present the information in terms of trees planted, kilometers of air travel, and daily emissions of an average EU citizen to facilitate the comprehension of the indicated amount of CO_2 emissions savings. These figures are in units that most respondents know from personal experience and can directly relate to. Also, to prevent experimenter demand effects—that is, "changes in behavior by experimental subjects due to cues about what constitutes appropriate behavior"

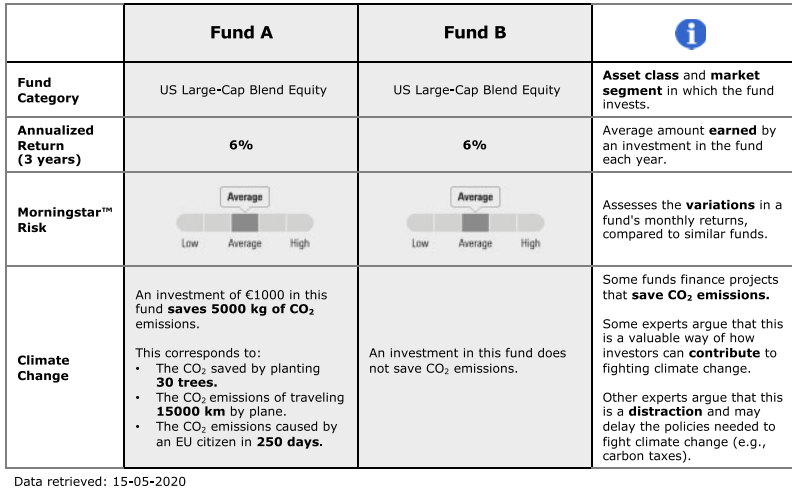


Figure 2
Investment information in the main experiment
This figure provides a screenshot of the information the investors participating in our main experiment receive on the two investments if they are assigned to the HIGHIMPACT treatment. The investment information investors in the LOWIMPACT treatment receive is shown in Figure A.3.

(Zizzo 2010, p. 75)—in addition to using financial incentives, we use an ambiguous framing of sustainable investing: in the information column on the right of Figure 2, we provide investors with arguments both for and against investing sustainably being socially desirable.

We randomly assign investors to one of two different treatments, HIGHIMPACT and LOWIMPACT. In HIGHIMPACT, the sustainable investment saves 5 tons of CO_2, whereas in LOWIMPACT it saves 0.5 tons of CO_2, 10 times less. To avoid ordering effects, we randomize whether the sustainable investment option is displayed on the screen's left or its right side. However, we do not find any significant ordering effects in our results (see Table A.1).

To guarantee that all relevant information is salient when the valuation decision is made, investors again need to participate in a brief quiz on the past performance, the risk level, and the impact of the two investments (see Figure A.3 for a screenshot). Investors who twice fail to answer the quiz questions correctly receive the correct answers and a short explanation.[7]

1.3 Investment decisions and WTP elicitation

In the third step, we elicit investors' WTP for the sustainable investment option. As the direct statement of a precise WTP is cognitively demanding for respondents and subject to noisy answers and outliers, we ask investors to

[7] Excluding investors who fail twice in at least one of these quizzes does not substantially affect our results, as shown in Table A.2.

make binary choices instead, which is the method most frequently used in the judgment and decision-making literature to measure (risk) preferences (Holt and Laury 2014). Our respondents repeatedly choose between the sustainable investment and the conventional investment, which are neutrally labeled as investments A and B. For each investment we indicate a one-time, upfront fee, which we vary between consecutive choices depending on respondents' answers. Both investment options start with the same fee, €10. If a participant prefers investment A, we increase the fee for investment A by €40 and ask again. Using the bisection method, also called the midweight method, we iteratively adjust the fee to elicit an investor's WTP (see, e.g., Abdellaoui 2000; van de Kuilen and Wakker 2011). We provide an illustrative example in Figure A.4, and additional details on the implementation of the bisection method in Section A.1. Using this method, we can determine the investors' WTP with a precision of €1.25 for the €1,000 investment through a series of seven choices.[8] To guarantee that our method yields each investor's true WTP, after the investment decisions have been made we ask respondents to confirm whether the elicited WTP actually reflects their true preferences. Respondents who do not agree with the elicited WTP are asked to repeat the procedure once, if they wish. We exclude investors who disagree with the elicited WTP and are unwilling to repeat the investment decisions.

1.4 Postexperiment survey

After the respondents make their investment decisions, they are asked to fill out a survey, which serves two purposes. First, we run a manipulation test to check whether investors understood and remember the investment information provided. We ask investors 1) to recall which investment had a higher impact and 2) how much impact the sustainable investment had exactly. Second, we ask questions about investors' financial expectations with regard to the investments, the feelings they associate with their choices, their perception of the impact of the sustainable investment, their individual preferences, as well as their demographic characteristics. Table A.4 summarizes all variables elicited in the postexperiment survey.

1.5 Participants

We conduct our main experiment with a sample of experienced private investors. We recruit them from the members of a Dutch investor protection interest group with some 40,000 members.[9] The main activities of this group are the provision of independent information for investors and the coordination of

[8] Some investors show a censored WTP for the sustainable investment: they do not deviate from the initially preferred investment in any of their seven choices. In this case, we ask these investors to directly state their WTP for the sustainable investment. Table A.3 shows our main results excluding all investors with censored WTP; the exclusion of these investors does not have a material effect on our results.

[9] Vereniging van Effectenbezitters (Association of Stockholders).

Table 1
Preferences and demographics for the private investors, by treatment

	Mean Values		Mann–Whitney U Test
	LowImpact (n=97)	HighImpact (n=99)	(HighImpact = LowImpact)
Risk preferences [0,10]	6.938	7.000	p=.307
Time preferences [0,10]	7.361	7.485	p=.389
Altruism [0,10]	6.588	6.455	p=.389
Climate awareness [0,10]	7.423	7.677	p=.694
Female [0,10]	0.103	0.131	p=.540
Age	61.660	61.495	p=.700
Income	€60,000–€74,999	€60,000–€74,999	p=.842
Net worth	€200,000–€499,999	€200,000–€499,999	p=.887
Highest education	Bachelor's degree	Bachelor's degree	p=.765
Investment knowledge [0,10]	6.318	6.234	p=.661

This table presents the preferences and demographic variables of the sample of private investors in our main experiment, by impact treatment. Time preferences, risk preferences, and altruism are measured on a 10-point scale using an experimentally validated survey module introduced by Falk et al. (2016). In order to improve readability, we transform other variables to a scale from 0 to 10. The first two columns report mean values of the variables, by impact treatment; the third column reports p-values of a Mann–Whitney U test, testing for differences between the two treatments.

lawsuits that aim to obtain compensation for groups of aggrieved shareholders. Its members hence have substantial experience of, and interest in, making investment decisions, which was our intent with this participant pool. For our experiments, we were able to recruit 527 participants via a general and regular newsletter to the members who subscribed to it. Of these participants, 219 take part in our main experiment; the remaining participants take part in two extensions of the main experiment. Table 1 shows the demographic characteristics and individual preferences of our sample of private investors in the main experiment. On average, our respondents are older and wealthier, and as a group have a higher share of males than that of the Dutch population. Both treatment groups, HighImpact and LowImpact, are well balanced in terms of demographic variables and individual preferences. We refer to these participants as private investors in the following.

1.6 Data processing

In our main experiment, we exclude from our analysis six investors who do not agree with the statement "Climate change is a serious problem that needs to be solved"—that is to say, investors who state an agreement level of 3 or less on a scale of 1 to 7. CO_2 savings are an inappropriate measure of impact for these investors, and we cannot detect how their WTP for sustainable investments relates to impact. We exclude a further 17 investors who explicitly disagree with the elicited WTP and are unwilling to repeat the investment decisions, as previously described.[10] This results in a final sample of 196 investors.

[10] We include 19 investors who disagree with the elicited WTP but are willing to repeat the investment decisions. We use the WTP calculated based on these repeated decisions for these investors.

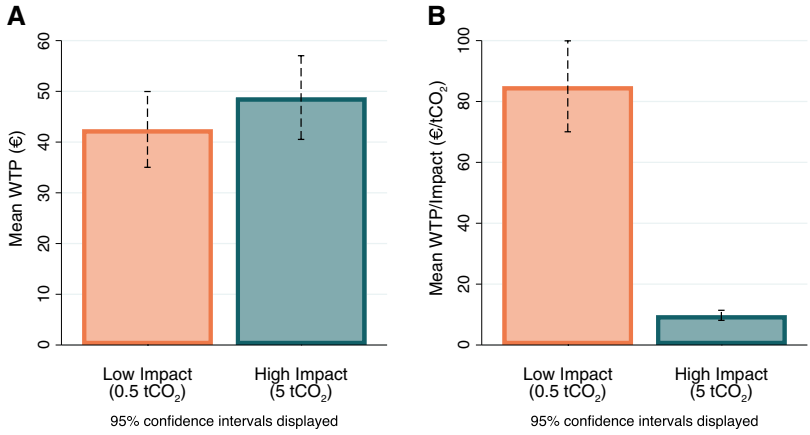


Figure 3
Response of private investors' WTP to the impact of sustainable investments.
This figure shows the results of our main experiment, which investigates how private investors' WTP for sustainable investments responds to the impact of these investments. Panel A: mean absolute WTP for the sustainable investment, by impact treatment. Panel B: mean relative WTP for the sustainable investment per ton of CO_2 saved, by impact treatment.

In accordance with our preregistered procedure, we winsorize all WTP values at the 5% and 95% levels to reduce the influence of extreme values.

2. Does WTP for Sustainable Investments Scale with Impact?

We observe that investors are willing to pay for investments with impact. Of all investors, 93% prefer the sustainable option when fees are equal in the two funds. Pooling investors in the LowImpact and the HighImpact treatment, the average WTP for the sustainable investment is €45.67 for a €1,000 investment. This substantial WTP is an important baseline finding for our investigation.

On this basis, we turn to our main question: Do investors have a higher WTP for investments with a higher impact? When contrasting the treatments, we find that the level of impact of sustainable investments does not significantly affect investors' WTP. Panel A of Figure 3 shows that investors' average WTP for the sustainable investment is €42.49 in the LowImpact and €48.78 in the HighImpact treatment. While investors do pay more in the HighImpact treatment, this difference is not significant ($p = .363$, Mann–Whitney U test). There is also no significant difference between the two treatments regarding the share of investors who prefer the sustainable investment ($p = .798$, Mann–Whitney U test). Also, a visual inspection of the distributions of private investors' WTP reveals no systematic difference between the treatments, as shown in Figure A.6, panel A.

Panel B of Figure 3 shows that the observed insensitivity to impact leads to a substantial inconsistency in investors' WTP per unit of impact. Investors are

Table 2
Results of the main experiment

	Mean Values		Mann–Whitney U Test
	LowImpact ($n=97$)	HighImpact ($n=99$)	(HighImpact = LowImpact)
Experimental Results			
WTP (€)	42.49	48.78	$p=.363$
WTP/Impact (€/tCO_2)	81.25	8.38	$p<.001$
Sustainable investment preference (%)	93.81	92.93	$p>.999$
Postexperiment Survey Results			
Risk expectations [−10,10]	−0.526	−0.051	$p=.382$
Return expectations [−10,10]	−0.312	−0.707	$p=.348$
Positive emotions [−10,10]	6.134	6.465	$p=.121$
Perceived investment impact [−10,10]	4.089	5.488	$p=.003$
General relevance impact [−10,10]	3.643	4.276	$p=.142$
General relevance impact level [−10,10]	2.474	2.896	$p=.457$
Estimated cost of saving CO_2 (€/tCO_2)	94.55	102.43	$p=.658$

This table presents the results of the main experiment. First, it shows private investor's absolute and relative WTP for the sustainable investment, as well as the share of investors who prefer the sustainable investment when fees are equal. The WTP is elicited using the experimental procedure described in Section 1. Second, it shows the results of the postexperiment survey. In order to improve readability, we transform variables from the postexperiment survey to a scale from −10 to 10. For risk expectations, return expectations, and positive emotions, positive values indicate that investors have a more favorable view of the sustainable investment; negative ones that they have a more favorable view of the conventional investment. The first two columns report mean values of the variables, by impact treatment; the third column reports p-values of a Mann–Whitney U test, testing for differences between the two treatments. Detailed descriptions of the variables can be found in Table A.4.

willing to pay significantly more per tCO_2 saved in the LowImpact treatment than in the HighImpact treatment ($p<.001$, Mann–Whitney U test). We observe a difference in the average WTP per ton of CO_2 saved of a factor of 9.7.

The postexperiment survey provides evidence that our results are not driven by differences in risk and return expectations for the sustainable investment between the HighImpact treatment and the LowImpact treatment. Such differences might conceal the influence of the investments' impact. As shown in Table 2, neither risk expectations nor return expectations differ significantly between the HighImpact and LowImpact treatments. Further, as shown in Table A.5, neither investors' risk expectations nor their return expectations correlate significantly with their WTP for the sustainable investment, and the effect of investments' impact on investors' WTP remains insignificant when we control for risk and return expectations.

Summing up, we find that, while investors have a substantial WTP for an investment with impact, they do not pay significantly more for more impact, even when the impact is increased by a factor of 10. This suggests that investors are quite willing to pay for sustainable investments, yet they have little regard for the amount of impact that such investments offer.

2.1 Robustness checks

We corroborate this main finding in several robustness checks. First, we confirm that our results are not driven by a lack either of the salience or of the

comprehensibility of the impact information provided. Second, we replicate the experiment with a focus on past financial performance, ensuring that our elicitation method can detect sensitivity when it is present. Third, we replicate the main experiment with a set of MTurk samples and test eight different variations of our elicitation method to address potential concerns that our results are an artifact of this method. Finally, we provide evidence that the COVID-19 crisis is unlikely to have affected our main results.

2.1.1 Is the impact information salient and comprehensible? Relying on the postexperiment survey, we examine whether the impact information is salient and comprehensible. If investors were not sufficiently attentive to the information provided, or if investors were unable to evaluate the impact information, this may explain the observed insensitivity.

First, we find that the impact information provided to the investors was salient during the investment decisions. Once the investment decisions had been made, we asked all participants if they could remember the impact information. We find that 99% could correctly identify the sustainable investment, and that 95% could, in a free text field, accurately reproduce its exact level of impact in tCO_2.

Second, we observe that, on average, investors made a realistic estimate of the value of saving a ton of CO_2 emissions. In the postexperiment survey, we ask investors for an estimate of the price of saving a ton of CO_2 emissions. On average, the investors' estimate of CO_2-saving costs is €98.55 per ton, with no significant difference between the HIGHIMPACT and LOWIMPACT treatments ($p = .658$, Mann–Whitney U test, 95% confidence interval: €77.08–€120.02). This is higher than the CO_2 prices in the European Union Emissions Trading System during our data collection period, which fluctuated roughly between €25 and €30. The values stated by investors do, however, correspond relatively well to estimates of the cost society incurs from carbon emissions. Based on a survey of a broad panel of climate scientists and economists, Pindyck (2019) estimates that the social cost of emitting a ton of CO_2 lies between $80 and $200. Besides the fact that we translate the CO_2 savings of the investments into more intuitively comprehensive units, this finding indicates that the information provided enables investors to evaluate the level of impact of the investments.

Finally, our results suggest that investors expect slightly different levels of impact conditional on the treatment. We asked investors whether they thought the sustainable investment makes a meaningful contribution to mitigating climate change ("Perceived investment impact" in Table 2). This variable is significantly higher in the HIGHIMPACT treatment than in the LOWIMPACT treatment ($p = .003$, Mann–Whitney U test). This suggests that investors seem to understand, at least in an ordinal sense, that the HIGHIMPACT treatment offers a higher impact.

2.1.2 Are our findings aligned with what investors say is important to them? We compare investors' WTP for impact with their statements about

the importance of impact. In the postexperiment survey, we ask two questions to this end. First, we ask investors how important it is to them that their investments contribute to halting climate change ("General relevance impact" in Table 2). Second, we ask investors how important it is to them *how much* their investments contribute to halting climate change ("General relevance impact level" in Table 2).

Investors assign importance to an investment having an impact and to how much impact an investment has ($p < .001$, Wilcoxon signed-rank test). At the same time, investors assign higher importance to the question of *whether* their investments contribute to climate change mitigation than to the question of *how much* their investments contribute to climate change mitigation ($p < .001$, Wilcoxon signed-rank test). These answers align with the fact that we find a WTP for investments with some impact but do not detect any significant differences in WTP between different levels of impact.

2.1.3 Can our elicitation method detect sensitivity when it is present? While so far only little is known about investors' sensitivity to impact, there is clear empirical evidence that investors are sensitive to mutual funds' past performance (e.g., Ivković and Weisbenner, 2009). Therefore, we apply our experimental procedure to measure investors' WTP for past financial performance. Concretely, we vary the investments' past performance, rather than their impact, between the two treatments. In each treatment, the baseline option has a past performance of 5% per year. In the HighReturn treatment, the second investment outperforms the baseline by 5% (i.e., a total performance of 10% per year). In the LowReturn treatment, the second investment outperforms the baseline by only 0.5% (i.e., a total performance of 5.5% per year). Hence, in accordance with the main experiment, outperformance differs by a factor of 10 between the two treatments. We do not provide information on impact in this setup. Using the same recruitment campaign as for the main experiment, we have a sample of 89 private investors who we randomly assign to this robustness check.[11] Using the same method as in our main experiment, we measure investors' relative WTP for the outperforming investment in terms of additional fees they are willing to pay.

We observe that investors' WTP responds strongly to the past performance of investment options. Figure 4 shows that investors are willing to pay significantly more in the HighReturn treatment than in the LowReturn treatment (panel A), and have a consistent WTP per unit of improved past performance across the treatments (panel B). The average WTP for the outperforming investment is a factor of 9.5 higher in HighReturn compared to investing in LowReturn: €121.22 versus only €12.82 ($p < .001$, Mann–Whitney U test). This difference is almost exactly proportional to the outperformance, which differs by a factor

[11] We exclude seven investors who explicitly disagree with the detected WTP and are unwilling to repeat the investment decisions.

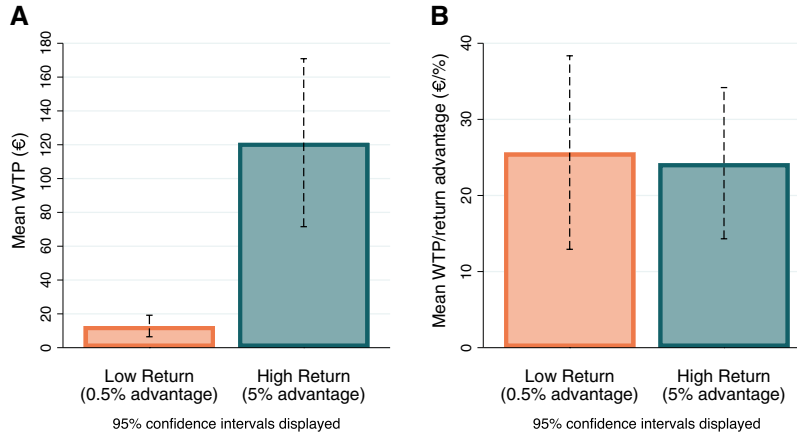


Figure 4
Response of private investors' WTP to the level of outperformance of investments.
This figure shows the results of an extension of our main experiment, investigating how private investors' WTP responds to the level of outperformance of investments. Panel A: mean absolute WTP for the investment with a higher past return, by treatment. Panel B: mean relative WTP for the investment with a higher past return, by treatment, expressed relative to outperformance in percentage points.

of 10 between the treatments. Based on these results, we conclude that our experimental design can detect investors' sensitivity when it is present.

2.1.4 Do our results replicate with different elicitation methods?

Potentially, our findings are affected by how we elicit investors' WTP. To explore the effect of our elicitation method on our results, we replicate our main experiment with a sample of 2,800 participants recruited through the crowd-working platform Amazon Mechanical Turk, implementing an exact replication as well as eight variations of our original setup. An overview of the results is shown in Table 3 and Figure 5. In short, the replications yield two main findings: First, the variations of the elicitation method change the average level of investors' WTP for the sustainable investment by up to $\pm$ 30%. Second, none of the variations produces a significant difference in WTP between the HighImpact and the LowImpact treatments.

We preregistered all replications. To ensure a high quality of observations from MTurk participants,[12] we used a vetted panel of reliable MTurkers provided by CloudResearch, we only accepted participants with outstanding reputation, we added an additional attention check, and we applied more stringent screening criteria with regard to our comprehension questions and the speed with which the experiment was conducted. We kept the variable incentive system constant for all replications, implementing a $1,000 investment for 10

[12] While data quality can be a concern with participants sourced on MTurk, experimental results have been shown to be reliable when appropriate quality control measures are applied (Arechar, Gächter, and Molleman 2018; Aguinis, Villamor, and Ramani 2021)

Table 3
Replications of the main experiment

Replication	N	Mean WTP ($)			Mann–Whitney U Test
		POOLED	LOWIMPACT	HIGHIMPACT	(HIGHIMPACT = LOWIMPACT)
R1: MTurk Baseline	1,000	43.91	42.97	44.93	$p = .412$
R2: Randomized Risk & Return	400	32.25	29.26	35.27	$p = .263$
R3: Direct Ask (No Anchor)	200	53.46	57.03	49.66	$p = .651$
R4: Direct Ask (Low Anchor)	200	33.89	34.60	33.19	$p = .093$
R5: Direct Ask (High Anchor)	200	37.06	35.45	38.71	$p = .321$
R6: No Neutral Choice	200	47.76	49.11	46.58	$p = .506$
R7: Upward Elicitation	200	33.02	35.35	30.87	$p = .611$
R8: Lower Scale	200	38.84	39.75	37.92	$p = .768$
R9: Higher Scale	200	55.22	48.30	63.18	$p = .084$

This table shows the key results of nine replications of the main experiment with MTurk samples. The table reports the number of observations, the pooled mean WTP across treatments, and the mean WTP per treatment. Further, it reports p-values of a Mann–Whitney U test, testing for differences in WTP between the treatments. "R1: MTurk Baseline" replicates our main experiment using the original procedure. "R2: Randomized Risk & Return" introduces additional variation regarding the risk and return of the sustainable investment. In "R3: Direct Ask (No Anchor)," we directly ask for investors' WTP for the sustainable investment instead of using a discrete choice design. "R4: Direct Ask (Low Anchor)" and "R5: Direct Ask (High Anchor)" provide different anchors regarding investment fees before the direct ask. "R6: No Neutral Choice" presents investors with an initial choice for which the sustainable investment already has a higher fee. "R7: Upward Elicitation", varies the bisection method: we start with a small fee increase and double the difference in fees in subsequent choices. "R8: Lower Scale" and "R9: Higher Scale" keep the original bisection method but decrease/increase the scale of fee differences by 50%. The detailed experimental procedures and preregistrations for these replications can be found in the appendix under Section A.2

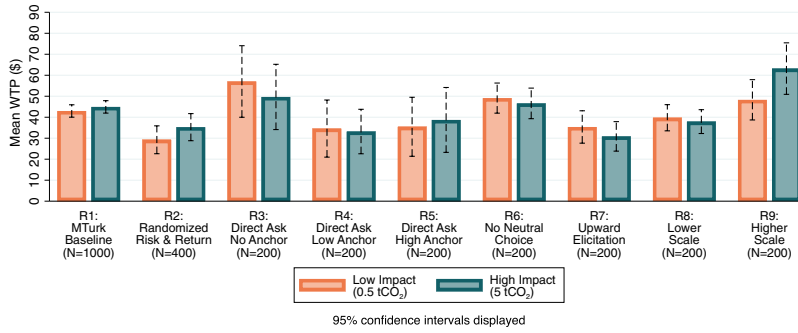


Figure 5
Summary of MTurk results
This figure shows the key results of nine replications of the main experiment with MTurk samples. The bars show the mean WTP per treatment and the error bars denote 5% and 95% confidence intervals. "R1: MTurk Baseline" replicates our main experiment using the original procedure. "R2: Randomized Risk & Return" introduces additional variation regarding the risk and return of the sustainable investment. In "R3: Direct Ask (No Anchor)," we directly ask for investors' WTP for the sustainable investment instead of using a discrete choice design. "R4: Direct Ask (Low Anchor)" and "R5: Direct Ask (High Anchor)" provide different anchors regarding investment fees before the direct ask. "R6: No Neutral Choice" presents investors with an initial choice for which the sustainable investment already has a higher fee. "R7: Upward Elicitation" varies the bisection method: we start with a small fee increase and double the difference in fees in subsequent choices. "R8: Lower Scale" and "R9: Higher Scale" keep the original bisection method but decrease/increase the scale of fee differences by 50%. The detailed experimental procedures and preregistrations for these replications can be found in the appendix under Section A.2.

randomly selected MTurk participants, based on their choices. The detailed experimental procedure of the replications and preregistration links can be found in the appendix under Section A.2.

First, we show that our main results replicate well with MTurk participants and that our findings are unlikely to be driven by a lack of statistical power. In the experiment "R1: MTurk Baseline," we replicate our main experiment with a sample of 1,000 MTurk participants, that is, a more than fivefold increase compared to the original sample size. The results of this replication are very similar to those of the main experiment. The pooled average WTP for the sustainable fund of \$43.91 we obtain for our MTurk sample is close to the original figure of €45.67 for the private investor sample. MTurk participants pay, on average, \$1.96 more in the HighImpact treatment, but the difference between treatments is insignificant ($p = .412$, Mann–Whitney U test). This indicates that our initial finding of a non-significant difference is not driven by a lack of statistical power. We also find that the quality of MTurk responses is high: 99% correctly remembered which investment had more impact, and 97% accurately reproduced the impact figures in the postexperiment survey. Taken together, these results indicate that our MTurk samples are a suitable reference to explore variations of the elicitation method in the following MTurk experiments.

Second, in the variation "R2: Randomized Risk & Return," we explore potential experimenter demand effects that may be induced by our design choices, concluding that such effects are unlikely to affect our findings substantially. In our main experiment, the two investment options differ only in terms of their impact, while the information on risk and return is identical. While we take several measures to reduce experimenter demand, participants might infer from this design aspect what we consider "appropriate" choices, which lead to increased WTP estimates. To address this concern, we introduce additional variation in the investments' risk and return characteristics. While we keep the risk and return characteristics of the conventional investment constant, we interact the two impact treatments with two treatment conditions regarding the risk and return of the sustainable investments: In the first condition, the sustainable investment features a lower past return and a better risk profile; in the second condition, it features a higher return and an inferior risk profile. As a result, there are four treatment conditions, and in each condition the sustainable investment differs from the baseline fund in all three dimensions (return, risk, and impact). This setup should make it harder for participants to infer what "appropriate" answers could be.

The average WTP for the sustainable investment across all treatments in this replication is \$32.25, which is significantly lower than in the baseline MTurk replication ($p = .001$, Mann–Whitney U test). Also, we detect a difference in WTP between the LowImpact treatment (\$29.26) and the HighImpact treatment (\$35.27) similar to the one in our main experiment. However, also in this specification, the difference is not significant ($p = .263$, Mann–Whitney

U test). Interestingly, participants report a lower level of experienced positive emotions derived for selecting this investment ($p < .001$, Mann–Whitney U test) compared to the baseline MTurk replication, which might explain why the pooled WTP is lower in this specification. We conclude that, while we cannot rule out that the focus on impact differences in our original procedure leads to somewhat inflated WTP estimates, our results hold in a setting in which this focus is substantially reduced.

Third, we explore how the specifications of the method we use to elicit investors' WTP affect our results and conclude that our findings essentially replicate with a broad set of alternative specifications. While, in three variations, participants' average WTP for the sustainable investment is slightly higher in the HighImpact treatment than in the LowImpact treatment, it is somewhat lower in four variations, so that there is no clear tendency. None of the differences between the two impact treatments are significant.

In a first set of variations, we omit the bisection method that confronts investors with consecutive binary choices and instead directly ask for investors' WTP for the sustainable investment. We implement three versions of this: In "R3: Direct Ask (No Anchor)," we do not provide any anchoring information on costs associated with investments; in "R4: Direct Ask (Low Anchor)," we provide the information that a passively managed fund may charge an annual fee of 0.1% per year; and in "R5: Direct Ask (High Anchor)," we provide the information that an actively managed fund may charge a 1% fee per year. The direct ask elicitation replication without anchor produces a significantly higher pooled WTP estimate than the baseline MTurk replication ($53.46, $p = .013$, Mann–Whitney U test). Conversely, the average pooled WTP is significantly lower for the replications with lower ($33.89) and higher anchors ($37.65) than in the MTurk baseline replication (in both cases $p < .001$, Mann–Whitney U test). In none of the direct ask variations do we find a significant difference in investors' WTP between the LowImpact and the HighImpact treatment.

In a second set of four replications, we vary several specifications of our discrete choice method. In the replication "R6: No Neutral Choice," we omit the first discrete choice of our original procedure, in which investors decide between the two investments with fees being equal. In the main experiment, most investors prefer the sustainable investment under these conditions. Thus, they may feel obliged to select the same investment in subsequent choices where it is more costly, which would inflate our WTP estimates. However, this seems unlikely, as removing the first choice leads to a slightly (but not significantly) higher pooled average WTP compared to the baseline replication ($47.76, $p = .181$, Mann–Whitney U test).

Next, we vary the bisection approach in the replication "R7: Upward Elicitation." In our original procedure, investors who have selected the sustainable investment in the first choice have to decide whether they are willing to pay $40 more in fees for this investment in the next choice. Again, investors may feel urged to accept this relatively high level of fees to remain

consistent with their initial choice. Thus, in this variation, we start with a much smaller fee increase of \$1.25. In subsequent choices, we double the difference in fees between the two investments until the participants deviate from their initial choice. Once the participant has switched, we iteratively reduce the fee difference as in the original method to obtain a more precise estimate. With this method, investors' WTP for the sustainable investment is significantly lower than in the baseline replication; however, it is still substantial (\$33.02, $p < .001$, Mann–Whitney U test).

Finally, in the replications "R8: Lower Scale" and "R9: Higher Scale," we keep the original bisection method but decrease or, respectively, increase the initial fee difference by 50%. Thus, in their second choice, participants have to decide whether they remain with their preferred investment if we increase its fee by \$20 ("R8: Lower Scale") or by \$60 ("R9: Higher Scale"). Decreasing our measurement scale by 50% leads to a slightly, but significantly, lower pooled average WTP for the sustainable investment than in the baseline replication (\$38.84, $p = .035$, Mann–Whitney U test). However, this difference is not proportional to the reduction in our measurement scale. When increasing our measurement scale by 50%, participants' pooled average WTP is substantially higher than in our baseline replication (\$55.22); however, the difference is not significant ($p = .112$, Mann–Whitney U test). Again, we do not find a significant difference between the LowImpact and the HighImpact treatments in any of the variations concerning the specifications of our discrete choice method.

The MTurk replications highlight that it is important to interpret the level of our WTP estimates in the context of our experimental design and data-processing procedures. The average WTP for the sustainable investment across both impact treatments of our main experiment is €45.67. This corresponds to a fee of roughly 4.5% of the investment, which might seem relatively high. More conservative elicitation methods yield pooled WTP estimates that correspond to fees of 3–3.5% of the investment ("R2: Randomized Risk & Return," "R4: Direct Ask (Low Anchor)," and "R7: Upward Elicitation"). Also, the data processing affects the level of our WTP estimates. In the preregistration, we commit to retain observations from participants with censored WTP,[13] and to mitigate the effect of outliers by winsorizing observations at the 5% and 95% level. If we deviate from the preregistered procedure and exclude investors with censored WTP, we obtain a pooled average WTP corresponding to a fee of roughly 3% for the main experiment (see A.3). In addition, it is relevant to consider that we measure WTP as an upfront fee for a 1-year €1,000 investment. Retail investors might be used to relatively high entry and exit

[13] Investors' WTP is censored if they do not deviate from the initially preferred investment in all seven investment choices. We cannot elicit WTP directly from the choices of these investors. We additionally ask these investors to state their WTP for the sustainable investment.

fees.[14] We conclude that our WTP estimates are somewhat sensitive to the elicitation method and data processing protocol, and that investors' WTP in the field might be lower than our estimates suggest. Importantly, however, our main result of insignificant treatment effects remains robust across different elicitation methods.

2.1.5 Are our results affected by the COVID-19 crisis? The COVID-19 crisis has been found to have affected investors' behavior in various ways (see, e.g., Ramelli and Wagner 2020). As our data collection took place in mid-2020, the pandemic may also have affected our results. To investigate the effect of the crisis, we make use of a preliminary version of our experiment, which we ran with 311 students at a large Dutch university in September 2019, well before the emergence of SARS-CoV-2. We repeated this experiment in September 2020, after the emergence of SARS-CoV-2, under the same conditions and with a corresponding sample of 243 students at the same university. We find no substantial differences between the results before and after the appearance of the virus. The detailed results can be found in Table A.6. In light of these results, it seems unlikely that COVID-19 had a relevant effect on our results.

2.2 Discussion

We find that investors are willing to pay to obtain a sustainable investment with some impact, but do not pay significantly more for more impact. We provide evidence that this finding is unlikely to result from methodological issues or from the fact that investors did not understand the impact information provided to them. We do not rule out completely that some investors have some degree of sensitivity to impact, but this seems to be a second- or third-order effect. What drives investors' WTP is their preference for a sustainable investment with some impact over a conventional investment with no impact.

This finding extends existing experimental studies in important ways. Prior studies have demonstrated that there is a WTP for sustainable investments (Riedl and Smeets 2017; Bauer, Ruof, and Smeets 2021; Barber, Morse, and Yasuda 2021), and that there are substantial financial flows toward sustainable investment funds (Hartzmark and Sussman 2019). However, these studies did not explore sensitivity to the level of impact. Our results suggest that investors have a positive WTP for sustainable investments as a category, but do not adjust their WTP according to the impact of these investments. This dovetails with an earlier finding of Riedl and Smeets (2017), who show that prosocial preferences explain whether investors invest in sustainable funds, but do not explain how much of their wealth they allocate to these funds.

[14] Entry and exit costs and brokerage charges can be substantial in the retail investing market. Khorana, Servaes, and Tufano (2009) estimate that the average shareholder cost for a 5-year investment period is 2.09% per year, of which 0.66% are entry and exit fees. Adjusted for a 1-year investment period, total costs would amount to 4.76%. According to more recent (2018) data, the average entry fee for equity funds in Europe was 3.65% (https://ec.europa.eu/info/publications/180425-retail-investment-products-distribution-systems_en).

Our findings are in contrast with traditional models of altruistic behavior. These models assume that individuals are consequentialists, in the sense that they contribute to public goods because they derive utility from the level of the public good, beyond the direct benefit they experience themselves from the good (Becker 1974; Eckel and Grossman 1996; Andreoni and Miller 2002). Such models, which are often labeled models of "pure" altruism, imply that the benefit an individual receives from performing a prosocial act depends on the act's impact on a public good. Thus, if prosocial investors were driven by pure altruism, we would expect their WTP for sustainable investments to increase with the impact of such investments. Yet, as our results show, even an increase in impact by a factor of 10 does not lead to a significant increase in investors' WTP. This observed behavior is not in line with a pure altruism decision model.[15]

Insensitivity to quantity has been demonstrated in other contexts, such as public good valuation and philanthropic donations. However, it is important to reconsider this phenomenon in the context of sustainable investing. First, the classic contingent valuation result by Desvousges et al. (1992) that individuals are willing to pay roughly the same amount to save 2,000, 20,000, or 200,000 birds relies on stated preferences. In a revealed preference setting where choices are consequential, individuals might evaluate their options more critically. Second, investment decisions tend to trigger a calculative decision mode, and thus public goods may be considered differently when they are part of an investment choice. Third, theorists currently model the behavior of sustainable investors as if they were consequentialists. Fourth, two contemporaneous working papers find that, within their experimental setups, respondents' WTP for sustainable investments scales with positive externalities (Bonnefon et al. 2022 Brodback, Günster, and Pouget 2021). Against this backdrop, our rejection of the consequentialist view of sustainable investors is an important piece of evidence. In the following, we explore reasons why investors' WTP does not significantly respond to the level of impact.

3. Exploring Reasons Why WTP for Sustainable Investments Does Not Scale with Impact

From the existing literature, we identify three potential explanations for the observed behavior. First, investors may lack the knowledge or experience

[15] Pure altruism could possibly explain the insensitivity we observe if the marginal societal utility of CO_2 emissions savings is strongly decreasing. This, however, seems unlikely in our setting. The pure altruist's marginal utility is proportional to the societal utility of an additional unit of impact. Thus, to explain the observed behavior one would have to assume that marginal societal utility is high up until 0.5 tons of CO_2 emissions savings and strongly decreases between 0.5 and 5 tons of CO_2 emissions savings. Such a case is hard to make from a societal perspective. A recent report by the United Nations Environment Programme (UNEP) concludes that annual global greenhouse gas emissions need to be reduced by 32 billion tCO_2 by 2030 if the internationally agreed upon target of limiting global warming to 1.5 degrees Celsius above preindustrial levels is to be reached (United Nations Environment Programme 2020). In light of this emission reduction gap, it seems highly unlikely that the marginal societal benefit decreases substantially for impact levels below 5 tCO_2, which is the highest impact level we use in our experiment.

necessary to evaluate impact information in an investment context. Second, investors may only be able to discriminate between products in terms of impact when they can directly compare several options. Third, investors' WTP may be driven by positive emotions, or, the warm glow that is associated with choosing a sustainable option, rather than by a calculative appraisal of that option's impact.

3.1 Do investors lack the necessary experience to evaluate the impact of investments?

Investors' lack of experience in evaluating impact information in an investment context could explain why their WTP for sustainable investments does not respond to the impact of those investments. We have already put several measures in place to ensure that investors understand the impact of the available investment options. Nevertheless, investors may still be unable to value this impact in the context of an investment decision. Even if they are able to differentiate between the impact of planting 3 and 30 trees, they may still be unable to evaluate whether planting 30 trees has an impact that is meaningful for a €1,000 investment. Research shows that the ability to perform such evaluations can increase with experience (Hsee and Zhang 2010).

To investigate the effect of experience on sensitivity to impact, we repeat our experiment with a unique sample of dedicated impact investors. We recruit this sample through different channels: first, from the alumni network of the University of Zurich's training programs for high-net-worth impact investors; second, via two associations of high-net-worth impact investors,[16] whose members come together to share knowledge and participate in events on impact investing; and third, from among the customers of five wealth advisors and private banks (those recruited were identified as experienced impact investors by their advisers).[17] In total, we recruited 125 impact investors through these channels.[18] The final sample comprises 118 investors as we excluded 1 who does not think that climate change is a serious problem and 6 who explicitly disagreed with the elicited WTP and were unwilling to repeat the investment decisions.

All of the investors in this sample have indicated their intention of exerting a positive impact with their investments, by taking courses, joining a network, or by instructing their advisors. Most have considerable experience in dealing with impact investments. While their individual levels of experience vary, we are confident that, on average, they have a much higher level of experience than the sample of private investors we recruited for our main experiment. If lacking

[16] Toniic and the NEXUS Working Group on Impact Investments.

[17] Credit Suisse, Bank Vontobel, Bank Julius Baer, Tiedemann Advisors, and VALUEworks.

[18] Five impact investors participated in the experiment shortly after the end of the sampling period specified in the preregistration. If we exactly follow the preregistration procedure, the sample is slightly smaller ($n = 120$). However, the results do not substantially differ from those obtained with the full sample (see Table A.7). As the sampling period does not seem to have an effect on our results, we analyze the entire sample.

Table 4
Preferences and demographics for the impact investors, by treatment

	Mean Values		Mann–Whitney U Test
	LowImpact (n=59)	HighImpact (n=59)	(HighImpact = LowImpact)
Risk preferences [0,10]	7.169	6.898	p=.521
Time preferences [0,10]	8.508	8.068	p=.119
Altruism [0,10]	7.763	7.169	p=.131
Climate awareness [0,10]	9.096	8.983	p=.814
Female [0,10]	0.356	0.407	p=.705
Age	41.424	38.966	p=.456
Income	€125,000–€149,999	€150,000–€174,999	p=.543
Net worth	€1M–€9.9M	€1M–€9.9M	p=.931
Highest education	Master's degree	Master's degree	p=.828
Investment knowledge [0,10]	6.877	6.707	p=.650

This table presents the preferences and demographic variables of our sample of impact investors, by impact treatment. Time preferences, risk preferences, and altruism are measured on a 10-point scale using an experimentally validated survey module introduced by Falk et al. (2016). In order to improve readability, we transform other variables to a scale from 0 to 10. The first two columns report mean values of the variables, by impact treatment; the third column reports p-values of a Mann–Whitney U test, testing for differences between the two treatments.

the ability to evaluate impact information drives insensitivity to impact, we would expect these impact investors to be more sensitive to impact than are the private investors.

Table 4 presents the characteristics of the impact investor sample. Compared to the private investors in the main experiment, the impact investors have greater wealth, state a higher level of investment experience, and are younger, and the sample has a larger fraction of female investors. The median household net worth lies between €1 million and €10 million.[19] Regarding their preferences, the impact investors are more altruistic and have a stronger long-term orientation than our private investors. Again, both treatment groups, HighImpact and LowImpact, are well balanced in terms of demographic variables and individual preferences.

The results of our experiment with dedicated impact investors are overall very similar to those of the main experiment with private investors. We find that impact investors too have a positive WTP for sustainable investments. Of all our impact investors, 97% prefer the sustainable investment when fees are equal. Pooling investors in the LowImpact and the HighImpact treatments, the average WTP for the sustainable investment is €49.01 for a €1,000 investment. This is slightly more than the figure for the private investors; the difference, however, is not significant (p=.096, Mann–Whitney U test).[20]

Further, we find that for impact investors too the level of impact of sustainable investments does not significantly affect their WTP. Figure 6, panel A contrasts impact investors' WTP for sustainable investments between the LowImpact

[19] This figure may be an understatement in numerous cases as many of the impact investors are embedded in family structures that collectively own much more, often several billion euros.

[20] A similar reasoning to that above applies to the magnitude of the measured WTP.

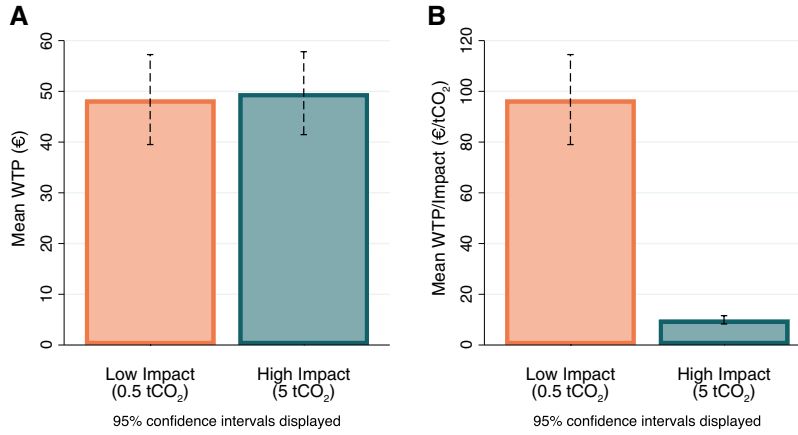


Figure 6
Response of impact investors' WTP to the impact of sustainable investments
This figure shows the results of our experiment investigating how impact investors' WTP for sustainable investments responds to the impact of these investments. Panel A: mean absolute WTP for the sustainable investment, by impact treatment. Panel B: mean relative WTP for the sustainable investment, per ton of CO_2 saved, by impact treatment.

and HIGHIMPACT treatments. There is no significant difference in the WTP for the sustainable investment between the treatments ($p = .767$, Mann–Whitney U test, Table 5). In the LOWIMPACT treatment, impact investors have an average WTP of €48.38 for an investment that saves 0.5 tCO_2, while in the HIGHIMPACT treatment the average WTP for an investment that saves 5 tCO_2 is €49.64. The distribution of impact investors' WTP for the sustainable investment can be found in Figure A.6, panel B.

Figure 6, panel B shows that for the impact investors too the WTP per unit of impact is inconsistent between the treatments. Impact investors are willing to pay significantly more per tCO_2 saved in the LOWIMPACT treatment than in the HIGHIMPACT treatment ($p < .001$, Mann–Whitney U test). The difference in the average WTP per ton of CO_2 is of a factor of 9.2.

As with the private investors, neither the risk expectations nor the return expectations of impact investors differ significantly between the HIGHIMPACT and LOWIMPACT treatments (Table 5). In comparison to the private investors, the impact investors have more positive expectations with regard to the financial performance of the sustainable investment. The impact investors expect the sustainable investment to have slightly better returns ($p = .047$, Mann–Whitney U test) and lower risk ($p = .006$, Mann–Whitney U test) than do the private investors.

Taken together, the results of our experiment with impact investors demonstrate that a lack of experience in evaluating the impact of investments is an unlikely explanation for the observed insensitivity to impact. We find that even dedicated and experienced impact investors do not respond, via their

Table 5
Results of the experiment with impact investors

	Mean Values		Mann–Whitney U Test
	LowImpact (n=59)	HighImpact (n=59)	(HighImpact = LowImpact)
Experimental Results			
WTP (€)	48.38	49.64	p=.767
WTP/Impact (€/tCO_2)	96.76	9.93	p<.001
Sustainable investment preference (%)	96.61	98.31	p>.999
Postexperiment Survey Results			
Risk expectations [−10,10]	0.678	0.593	p=.991
Return expectations [−10,10]	0.169	0.254	p=.952
Positive emotions [−10,10]	7.797	6.864	p=.209
Perceived investment impact [−10,10]	3.898	5.085	p=.314
General relevance impact [−10,10]	6.158	6.158	p=.820
General relevance impact level [−10,10]	5.763	4.746	p=.182
Estimated cost of saving CO_2 (€/tCO_2)	404.57	291.47	p=.258

This table presents the results of the experiment with impact investors. First, it shows impact investors' absolute and relative WTP for the sustainable investment, as well as the share of investors who prefer the sustainable investment when fees are equal. The WTP is elicited using the experimental procedure described in Section 1. Second, it shows the results of the postexperiment survey for the sample of impact investors. In order to improve readability, we transform variables from the postexperiment survey to a scale from −10 to 10. For risk expectations, return expectations, and positive emotions, positive values indicate that investors have a more favorable view of the sustainable investment, negative ones that they have a more favorable view of the conventional investment. The first two columns report mean values of the variables, by impact treatment; the third column reports p-values of a Mann–Whitney U test, testing for differences between the two treatments. Detailed descriptions of the variables can be found in Table A.4.

WTP, to different levels of investment impact. We therefore conclude that it is not a mere lack of experience that drives insensitivity to impact. This finding prefigures one important implication—namely, that investor training and the building up of experience may not be sufficient to address the problems that come with the observed insensitivity to impact.

3.2 Does comparability increase investors' sensitivity to impact?

The choice investors face in our main experiment corresponds to one that many retail investors face when their bank advisor offers them a binary choice between a conventional and a sustainable investment product. However, this may not necessarily correspond to the choice faced by more experienced investors, who can evaluate a broader set of investment options. As demonstrated by Hartzmark and Sussman (2019), these investors may consider information sources like the Morningstar "Globe" Rating, which allow them to compare a range of different investment options.

To investigate the importance of relative comparison, we run our experiment in a joint evaluation setup. Investors receive information on three investments: Fund A has zero impact, Fund B has a comparatively low level of impact, and Fund C has a considerably higher level of impact. Figure 7 illustrates the experimental procedure of our joint evaluation extension. We divide investors into two treatments, which we denote as the HighImpactRange treatment and the LowImpactRange treatment. In the HighImpactRange treatment, Fund B

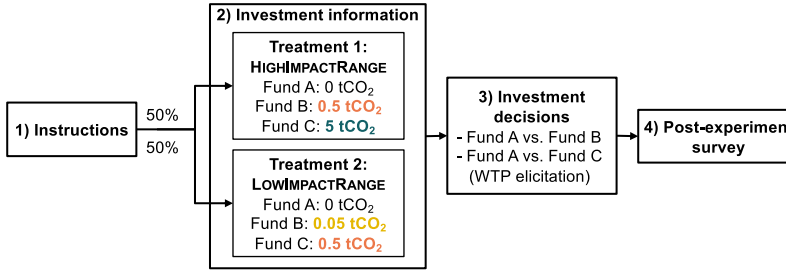


Figure 7
Experimental procedure of the joint evaluation extension
This figure provides an illustration of the experimental procedure we use in the joint evaluation extension of our main experiment.

	Fund A	Fund B	Fund C	
Fund Category	US Large-Cap Blend Equity	US Large-Cap Blend Equity	US Large-Cap Blend Equity	**Asset class** and **market segment** in which the fund invests.
Annualized Return (3 years)	**6%**	**6%**	**6%**	Average amount **earned** by an investment in the fund each year.
Morningstar™ Risk	Average Low Average High	Average Low Average High	Average Low Average High	Assesses the **variations** in a fund's monthly returns, compared to similar funds.
Climate Change	An investment into Fund A does not save CO_2 emissions.	An investment of €1000 in this fund **saves 500 kg of CO_2** emissions. This corresponds to: • The CO_2 saved by planting **3 trees.** • The CO_2 emissions of traveling **1500 km** by plane. • The CO_2 emissions caused by an EU citizen in **25 days.**	An investment of €1000 in this fund **saves 5000 kg of CO_2** emissions. This corresponds to: • The CO_2 saved by planting **30 trees.** • The CO_2 emissions of traveling **15000 km** by plane. • The CO_2 emissions caused by an EU citizen in **250 days.**	Some funds finance projects that **save CO_2 emissions.** Some experts argue that this is a valuable way of how investors can **contribute** to fighting climate change. Other experts argue that this is a **distraction** and may delay the policies needed to fight climate change (e.g., carbon taxes).

Data retrieved: 15-05-2020

Figure 8
Screenshot of the investment information in the joint evaluation extension
This figure provides an example of the information the investors receive with regard to the three investments in the joint evaluation extension of our experiment. The screenshot corresponds to the investment information investors in the HIGHIMPACTRANGE treatment receive. The investment information investors in the LOWIMPACTRANGE treatment receive can be found in Figure A.5.

saves 0.5 tCO_2 and Fund C saves 5 tCO_2. This corresponds to the impact values of the two treatments in our main experiment. In the LOWIMPACTRANGE treatment, Fund B only saves 0.05 tCO_2 and Fund C only saves 0.5 tCO_2. Again, for each treatment we translate this impact into more intuitively comprehensible units, as shown in Figure 8 for the treatment HIGHIMPACTRANGE.

For each investor, we assess the WTP for Fund B and for Fund C one after the other relative to Fund A, using the same procedure as in the main experiment. Hence, once one WTP is determined, the participant goes through the same mechanism again with the other sustainable investment in comparison to Fund A. During both WTP elicitations, investors always see the information on all

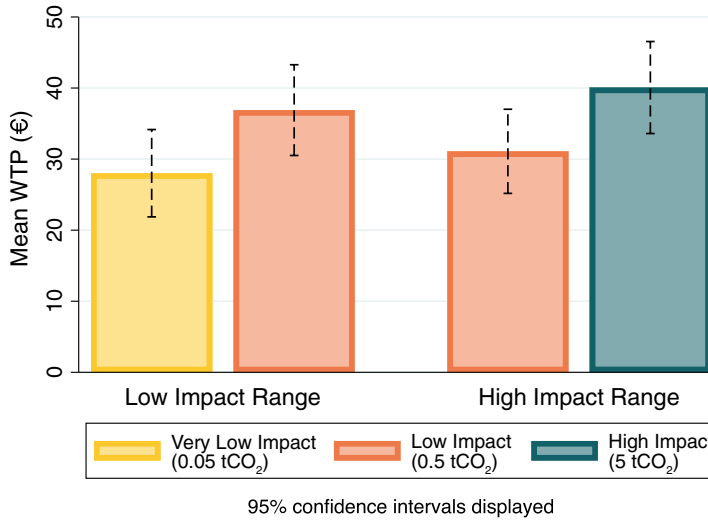


Figure 9
The effect of joint evaluation on private investors' response to impact
This figure shows the results of our experiment investigating how increased comparability affects the response of private investors' WTP to the impact of investments. The graph illustrates the mean WTP for the sustainable investments, by the investment's impact and treatment group.

three investments, as illustrated in Figure 8. The order in which WTP is elicited, first for Fund B or first for Fund C, is randomized.

We run the experiment with 219 additional private investors from the same recruitment campaign that we used for the main experiment. We exclude 11 investors who do not agree that climate change is a serious problem and 11 who explicitly disagree with the detected WTP and are unwilling to repeat the investment decisions. This results in a final sample of 197 investors.

The findings of the joint evaluation extension show that when comparing different sustainable investments, investors' WTP reacts to some extent to differences in impact (Figure 9). The results of the HIGHIMPACTRANGE treatment, in which investors see the same sustainable investments we use in our main experiment, provide evidence that increased comparability leads to a significant difference in WTP between the two sustainable investments ($p < .001$, Wilcoxon signed-rank test). The mean WTP is €31.09 for the sustainable investment that saves 0.5 tCO_2 and €40.07 for the sustainable investment that saves 5 tCO_2. While this indicates that investors respond to differences in impact in the case of directly and easily comparable options, their sensitivity remains limited: an increase in impact by a factor of 10 increases investors' WTP by only 28%. Thus, the value investors assign to a unit of impact remains inconsistent; there is a significant difference in the WTP per unit of impact between the two available sustainable investments ($p < .001$, Wilcoxon signed-rank test).

Further, the results show that even in the joint evaluation setup, investors' WTP does not respond to the absolute level of impact. The results in the LowImpactRange treatment are very similar to those observed in the HighImpactRange treatment: investors' WTP for the sustainable investment that saves 0.05 tCO_2 is €28.01; the figure is €36.89 for the investment that saves 0.5 tCO_2. Despite the tenfold difference in impact, we find no significant difference if we compare the WTP for the more impactful sustainable investment (Fund C) between the HighImpactRange and the LowImpactRange treatments ($p = .394$, Mann–Whitney U test). There is also no significant difference if we compare the WTP for the sustainable investment with a lower impact (Fund B) between the HighImpactRange and the LowImpactRange treatments ($p = .273$, Mann–Whitney U test).

On average, investors' WTP is lower in the joint evaluation extension than in our main experiment. The pooled average WTP over both sustainable investments in the HighImpactRange treatment of the joint evaluation experiment is €34.12. This is significantly lower than the pooled average over both impact treatments in the main experiment, in which the sustainable funds feature the same levels of impact ($p < .001$, Mann–Whitney U test).

The joint evaluation allows us to analyze within-subject sensitivity to the impact of investments, and we observe substantial heterogeneity in individual investors' sensitivity. We define the variable sensitivity as investors' WTP for Fund C (higher impact) divided by the WTP for Fund B (lower impact). Table 6 shows that about one-third of investors are entirely insensitive to investments' impact (labeled "Insensitive"). Even with a side-by-side comparison, these investors' WTP is equal for both sustainable investments, on average €34.5. We split those investors that are willing to pay more for more impact into two groups at their median sensitivity (labeled "low sensitivity" and "high sensitivity"). Investors in the "low sensitivity" group are, on average, willing to pay 1.3 times more for the more impactful investment. Their average WTP for the sustainable investments is higher than the one for the other investors (€46.5, $p < .001$, Mann–Whitney U test). In the "high sensitivity" group, investors' WTP is, on average, six times higher for an investment with 10 times more impact. Yet their average WTP for both sustainable investments is significantly lower than that of the other sensitivity groups (€21.2, $p = .005$, Mann–Whitney U test). Investors in the "high sensitivity" group have a higher level of education and earn more than the other investors ($p = .033$ and $p = .004$, Mann–Whitney U test). These findings suggest that the modest overall increase in sensitivity in the joint evaluation setup is driven by a subgroup of investors who are really sensitive to impact.

Taken together, the joint evaluation demonstrates that comparability creates some sensitivity to impact. However, even with options to compare, the average investors' WTP for sustainable investments is still far from proportional to the impact of these investments. First, within the two treatments, a clearly visible tenfold increase in impact leads to a WTP that is approximately 30% higher.

Table 6
Investors' level of sensitivity to the impact of investments

	N	Frequency	Mean WTP (€)			Mean Sensitivity
			FUND B	FUND C	AVERAGE	
Inverse sensitivity	9	5%	40.3	22.7	31.5	0.5
Insensitive	67	34%	34.5	34.5	34.5	1.0
Low sensitivity	61	31%	41.6	51.3	46.5	1.3
High sensitivity	60	30%	10.1	32.2	21.2	6.0

This table provides an overview of investors' level of sensitivity to impact in the joint evaluation experiment. We define the variable sensitivity as each investor's WTP for Fund C (higher impact) divided by the WTP for Fund B (lower impact). The values reported in the table pool observations from the HIGHIMPACTRANGE and the LOWIMPACTRANGE treatments. We divide investors into four groups with respect to their level of sensitivity. Investors with "Inverse sensitivity" pay less for the investment with a higher impact. "Insensitive" investors have the same WTP for both investments. The remaining investors pay more for the investment with higher impact and are split into two equally large groups ("low sensitivity" and "high sensitivity"). For each group, we report the number of observations, the mean WTP for Funds B and C individually and pooled, as well as the mean sensitivity.

Second, the choice set strongly influences investors' WTP per unit of impact. This indicates that investors evaluate the impact of investments relative to other available options. Further, the results show that, while comparability increases investor's sensitivity to impact, it also tends to diminish their baseline WTP for sustainable investments.

3.3 Is investors' valuation of impact driven by emotion rather than by calculation?

A third potential explanation for investors' insensitivity to impact is that their valuation is driven by the emotional experience of choosing the sustainable option rather than by a calculative appraisal of the impact of this choice. This idea is in accordance with models of warm glow, in which individuals' utility is unrelated to the level of the public good, being instead related to an emotional response that comes from the prosocial act itself (Andreoni 1989; Andreoni 1990). Further, Hsee and Rottenstreich (2004) argue that when individuals value a good's characteristic based on emotional perception rather than calculative appraisal, their WTP tends to be a step function of the characteristic. They show, for example, that the willingness to donate money to save pandas depends on the emotional importance of pandas in general, not on the number of pandas that will be saved. Applying this to our context, the emotional response to choosing a sustainable investment could explain the WTP for this sustainable investment.

Our postexperiment survey shows that choosing a sustainable investment feels good to investors (Table 7). We ask investors how good it feels to invest in the sustainable investment compared to in the non-sustainable one. Both private investors and impact investors report that it feels better to invest in the sustainable investment ($p < .001$, Mann–Whitney U test). The impact investors report a higher level of positive emotions as a consequence of choosing the sustainable investment than do the private investors ($p = .005$, Mann–Whitney U test). However, for both private investors and impact investors

Table 7
Positive emotions associated with investing sustainably

	N	Mean Positive Emotions [−10,10]		Mann–Whitney *U* Test
		LowImpact	HighImpact	(HighImpact = LowImpact)
Private investors	196	6.1	6.5	$p = .121$
Impact investors	118	7.8	6.9	$p = .209$
MTurks	1,000	6.5	6.5	$p = .710$

This table presents the self-stated emotions investors experience when choosing the sustainable investment, per treatment, and sample. Positive emotions are denoted on a scale of −10 to 10, where positive values indicate that it feels better for investors to choose the sustainable investment, and negative values that it feels better to choose the conventional investment. The first two columns report mean values of the self-stated positive emotions, by impact treatment; the third column reports *p*-values of a Mann–Whitney *U* test, testing for differences between the two treatments. "Private investors" is the sample from the main experiment, "Impact investors" is the sample of impact investors, and "MTurks" is the sample from the "R1: MTurk Baseline" replication.

the investment's impact does not affect these positive emotions. There is no significant difference in the reported positive emotions between HighImpact and LowImpact for both samples. This result also holds for the replication of our main experiment with a larger MTurk sample. If positive emotions drive investors' WTP for sustainable investments, it would explain why there is no difference in the WTP between the treatments.

Further, we show that investors' valuation of a unit of impact increases with positive emotions, but not with their cost estimate for a unit of impact. In Table 8, we regress investors' WTP per ton of CO_2 on the level of positive emotions investors experience and their estimate of the cost of saving a ton of CO_2. We find that for both private investors and impact investors, the WTP per ton of CO_2 significantly correlates with the level of reported positive emotions, but not with investors' estimate of the cost of saving CO_2. For our larger sample of MTurks in the replication of our main experiment, there is a significant correlation between participants' WTP per ton of CO_2 and their estimate of the cost of saving CO_2. However, the effect is relatively weak: A \$1 increase in the CO_2 cost estimate corresponds to less than a cent increase in the WTP per ton of CO_2. The significance does not persist if we control for demographics. Consistent with our previous findings, for all samples, the impact treatment has a highly significant effect on participants' valuation of a unit of impact. The regression results suggest that investors' valuation of impact is strongly influenced by emotions. A calculative appraisal, in which investors determine their WTP by estimating what might be an appropriate price for the impact that is offered to them, does not seem to play an important role.

An emotional valuation mode seems a reasonable explanation not only for the result of our main experiment but also for the results of the extensions. Regarding the extension with dedicated impact investors, experience can be expected to lead to an improved calculative valuation, as subjects have a better frame of reference through which to price impact. However, when investors' WTP depends mainly on their level of positive emotions, greater precision in estimating costs may be irrelevant. Regarding the role of comparability, "relative" emotional valuation could explain why in the joint evaluation

Table 8
Emotions, cost estimates, and investors' valuation of impact

	Private Investors		Impact Investors		MTurks	
	(1) WTP/tCO$_2$	(2) WTP/tCO$_2$	(3) WTP/tCO$_2$	(4) WTP/tCO$_2$	(5) WTP/tCO$_2$	(6) WTP/tCO$_2$
Positive emotions	4.341***	4.314***	3.408**	2.668*	4.826***	4.680***
	(0.968)	(0.957)	(1.229)	(1.326)	(0.384)	(0.388)
Estimated cost of saving 1 ton of CO$_2$	0.0318	0.0165	−0.0117	−0.00976	0.00232*	0.00197
	(0.0235)	(0.0237)	(0.00781)	(0.00831)	(0.00113)	(0.00114)
Impact treatment	−77.76***	−78.21***	−86.57***	−86.07***	−76.17***	−76.14***
	(7.132)	(6.998)	(8.919)	(9.295)	(2.919)	(2.916)
Demographics	No	Yes	No	Yes	No	Yes
Constant	56.20***	−63.67	74.91***	15.94	52.51***	83.37***
	(8.124)	(44.96)	(11.80)	(46.45)	(3.349)	(17.38)
Observations	195	195	117	117	1,000	1,000
R^2	0.416	0.486	0.492	0.524	0.464	0.474
F	45.40	10.53	36.54	9.534	287.0	74.12

* $p < .05$; ** $p < .01$; *** $p < .001$.
This table presents the results of an ordinary least squares (OLS) regression with WTP for the sustainable investment per unit of impact as the dependent variable. In all specifications, investors' level of self-stated positive emotions experienced when choosing the sustainable investment as well as their estimate of the cost of saving a ton of CO$_2$ are included as independent variables, together with an indicator variable taking the value of 0 for the LowImpact treatment and 1 for the HighImpact treatment. In addition, specifications (2) and (4) also include controls for investors' demographics, as described in detail in Table A.4. Specifications (1) and (2) report the results for our sample of private investors in the main experiment; specifications (3) and (4) report the results for our sample of impact investors and specifications (5) and (6) report the results for our sample of MTurks in the "R1: MTurk Baseline" replication of the main experiment. Standard errors are shown in parentheses.

extension investors respond to differences in impact but not to the absolute level of impact. In a direct comparison, it is obvious which option offers greater impact; thus, in a calculative valuation mode WTP for this better option is likely to be greater. This, however, may also be true in an emotional valuation mode. As argued by Hsee and Rottenstreich (2004), given a choice between saving two or saving three pandas, it is easy to see that the emotional return will be lower when saving only two. Along the same lines, Ferguson and Flynn (2016) propose a model of relative warm glow in which the warm glow individuals derive from choosing an option depends on how "good" this option is relative to other options in a given choice set. Such a relative emotional valuation is consistent with both findings of our comparative settings. Investors' WTP responds to impact within the choice set, but it remains constant when the choice set as a whole shifts to a different level of impact. When WTP is determined by emotions, these two observations are easily reconciled.

The level of positive emotions investors experience when selecting sustainable investments may depend on the choice environment. It appears that those choice settings that emphasize a more calculative valuation of impact reduce the positive emotions associated with choosing the sustainable option. Both our joint evaluation extension and the MTurk replication in which the investments differ in their risk and return ("R2: Randomized Risk & Return")

likely increase the focus on calculative valuation, and we observe lower average levels of WTP in these experiments. While we do not measure the positive emotions associated with selecting a sustainable choice in the joint evaluation extension, participants in the replication "R2: Randomized Risk & Return" report a significantly lower level of such emotions. Along similar lines, Karlan and Wood (2017) find that small-scale donors are less likely to give again if provided with quantitative information about a charity's effectiveness, presumably because it reduces the warm glow they derive from donating. Overall, the best explanation for our results throughout the paper remains that investors' WTP for sustainable investments is primarily driven by emotional experience, or warm glow.

4. Implications

Based on our findings, we argue that the average sustainable investor does not behave like a consequentialist, but rather like a warm-glow optimizer. Investors explicitly give up financial wealth to invest sustainably, but this WTP is driven by positive emotions regarding making an impact rather than by a calculative valuation of impact.

A key implication of this conclusion is that there may be incentives for greenwashing, or impact washing,[21] in the market for sustainable investment products. With an estimated size of USD 35 trillion assets under management (Global Sustainable Investment Alliance 2021), the market for sustainable investment products is already substantial, and it is set to grow further. If we accept the premise that sustainable investments may help to reduce negative externalities in the real economy, this rapid growth is promising. Our results, however, illustrate that investors pay for positive emotions regarding impact rather than for impact itself. Assuming impact is costly,[22] this creates an incentive for financial institutions to create products that optimize investors' emotions rather than impact.

The current market for sustainable investment products already reflects this focus on emotions. Most sustainable investment products cater to investors' desire to have an impact with their investing, at least in an implicit way. But it is often a vague promise, and few products underpin this promise with evidence. A review of 230 European investment funds with a focus on sustainability found that 52% made explicit environmental impact claims in their marketing

[21] (Busch et al., 2021, p. 2) define impact washing as "the dilution of the term impact investing using the term impact as a marketing tool to attract capital or boost reputations without actually focusing on material solutions to environmental and societal challenges."

[22] Several economic models imply that prosocial investors need to accept lower financial performance in order to have impact (Oehmke and Opp 2019; Heinkel, Kraus, and Zechner 2001; Pastor, Stambaugh, and Taylor 2021). Also, sustainable investments have additional data and expertise requirements that are likely to add to product costs and fees.

material, but none of the claims was substantiated.[23] Our study suggests that one key reason for that situation might be that investors are not demanding such evidence, because their preference is to make a choice that is emotionally rewarding. In turn, fund managers are not competing on impact and have little incentive to provide quantitative evidence of it. This is in line with the observation made by Hartzmark and Sussman (2019), who find that market-wide flows strongly increase when funds are clearly labeled as sustainable, while a fund's performance in more quantitative sustainability indicators has a negligible effect on fund flows. A further indication comes from Gibson et al. (Forthcoming), who find that the holdings of institutional investors that are committed to sustainability are only marginally more sustainable, especially in terms of environmental sustainability. These findings are consistent with a market for sustainable investing that is driven by a demand for positive feelings rather than a calculative valuation of impact.

Certainly, catering to investors' emotions is a business opportunity. Laudi, Smeets, and Weitzel (2021) provide experimental evidence that professional financial advisors actively exploit the sustainability preferences of their clients. In real settings, products may be marketed strategically in such a way that a sustainable product with relatively modest impact is presented as the sustainable option. This would enable financial institutions to collect investors' WTP for sustainable investments, while avoiding the costs of developing investment products with substantial positive impact. This creates the risk of a market for sustainable investment products that benefits investors in terms of warm glow and financial institutions in terms of fees, but falls well short of fulfilling its potential for solving important societal problems such as curbing carbon emissions to combat climate change.

Our findings also challenge an important assumption in the literature on asset pricing and sustainable finance. Several theoretical papers present models that explore prosocial investors' impact on asset prices (Oehmke and Opp 2019; Broccardo, Hart, and Zingales 2020; Fama and French 2007). These models suggest that prosocial investors shift asset prices, and reward firms that reduce negative externalities with a lower cost of capital. An important assumption in these models, embedded in the utility function of investors, is that prosocial investors optimize between the cost and the impact of sustainable investment opportunities.

We find that such an assumption may be problematic. We suggest that, rather than optimizing for impact, prosocial investors optimize for warm glow, and that in turn warm glow depends on the way in which investment options are presented and marketed to them. As a consequence, the predictions of these models may be overly optimistic regarding the effect of prosocial preferences on asset prices and externalities. Rather than rewarding firms that effectively

[23] See https://2degrees-investing.org/wp-content/uploads/2020/03/Report-EU-Retail-Funds-Environmental-Impact-Claims-Do-Not-Comply-with-Regulatory-Guidance.pdf.

reduce negative externalities, investors with prosocial preferences may be rewarding financial intermediaries that design and market products that offer warm glow. Exploring the consequences of warm-glow optimization for models that link prosocial preferences to asset prices is therefore an important avenue for future research.

Finally, our results have implications for the metrics that are used in sustainable investing. Problems of measurement are pervasive in sustainable finance. Claims that an investment is "sustainable" or the fact that it has one additional Morningstar "globe" are difficult to interpret even for experts. Confronted with this challenge, investors may appear insensitive to quantity simply because it is unclear to them what quantity means. Our study shows that scope insensitivity persists when impact is presented as a relatively well-defined quantity. Tons of CO_2 emissions are an established metric that investors can understand. And yet, investors remain insensitive to quantity when confronted with CO_2 emissions. This suggests that improved sustainability accounting, while desirable in many ways, will not help to realign WTP with impact, because valuation is being driven by emotions.

One potential solution to this problem is revealed by a comparison between our results and those of Bonnefon et al. (2022) and Brodback, Günster, and Pouget (2021). In their experiments, it appears that investors consistently value positive externalities if these are expressed in monetary terms. Of course, monetizing environmental and social impacts is difficult and the typical environmental social and governance (ESG) product is a long way from providing such a figure. Efforts to provide monetized impact measures are, however, underway.[24] Equipped with monetized impact measurements, investors may indeed one day behave as consequentialists who adjust their WTP to the level of impact that a product offers them.

Another way to realign emotional preferences with quantitative product characteristics is labels for sustainability funds. In Europe, a number of labeling initiatives to this end have already been established, and our results underscore the importance of such efforts.[25] In the United States, meanwhile, there are currently no established labels for sustainable investment products. Yet, there are also risks inherent in labeling. Our findings demonstrate that investors have a substantial WTP even for investments with a rather modest impact if these are perceived as more impactful than the alternative. It is thus important to ensure that labels actually align with socially desirable outcomes. So, for example, a product receiving a high score in a climate rating should respect limiting global warming to 1.5° Celsius above preindustrial levels—the international goal agreed upon at the Paris climate summit in 2015.

---

[24] See for example https://www.hbs.edu/impact-weighted-accounts/Pages/default.aspx.

[25] There are eight different sustainability labels provided by NGOs for funds in Europe; there is the Morningstar Sustainability Rating, described in Hartzmark and Sussman (2019); and there is an ongoing effort by the European Union to develop an "ecolabel" for sustainable investment products.

5. Conclusion

We present evidence that investors' willingness-to-pay for sustainable investments is largely independent of the impact of such investments. We arrive at this result for both experienced private investors and dedicated high-net-worth impact investors. The results replicate with a larger MTurk sample and for variations of the elicitation method. Being able to exclude a series of alternative explanations, we suggest that prosocial investors are best understood as warm-glow optimizers, who prefer investments that feel good, rather than as consequentialists, who derive utility from optimizing their impact.

Our findings have important implications for modeling investors' prosocial preferences in asset pricing and for policy makers who want to harness the growing demand for sustainable investments in order to support efforts to achieve societal goals. Current theoretical models routinely assume prosocial investors to be consequentialists. Incorporating the importance of warm glow for decision-making might affect these models' conclusions. For policy makers, our findings indicate a risk of greenwashing by providers of financial products and a potential equilibrium of "light green" products. This is particularly relevant against the backdrop of achieving internationally agreed upon sustainability goals, and given the fact that we observe our results even for experienced and dedicated impact investors.

A. Appendix

A.1 Detailed procedure: WTP elicitation and incentivization

A.1.1 WTP elicitation. We elicit investors' WTP for the sustainable investment through a series of seven binary choices between Fund A and Fund B. In the first choice, both funds feature a one-time, upfront fee of €10, which will be deducted from the €1,000 investment. This first choice reveals which investment option investors prefer if fees are equal. By way of explanation, assume an investor initially chooses Fund A, and that this is the sustainable option.[26] In the next step, we add €40 to the fee for Fund A, so that the investor now has the choice between Fund A with a fee of €50, and Fund B with a fee of €10. If an investor switches to Fund B under these conditions, we deduct €20 from the fee for Fund A for the following choice. In this case, the investor now faces a choice between Fund A with a fee of €30, and Fund B with a fee of €10. If, rather, an investor still prefers Fund A, we add another €20 to the fee for Fund A for the following choice, so that the investor has the choice between Fund A with a fee of €70, and Fund B with a fee of €10. We repeat this procedure over four additional choices, dividing the amount we deduct from or add to the fee for Fund A by a factor of two in each consecutive round. During each choice, all information on the two investments is visible to the investors.

After these seven choices, we calculate the midpoint between the highest fee the investor is willing to accept for Fund A and the lowest fee for Fund A at which the investor decided to switch to Fund B. We calculate the investor's WTP for Fund A as this midpoint minus the baseline fee of €10. For investors who prefer Fund B in the initial choice, we proceed accordingly by varying the fees for Fund B. If Fund B is the investment without impact, we elicit the WTP for this investment in the same way as described above, but we use the negative of this value as the investor's WTP

[26] Note that we randomize whether Fund A or Fund B corresponds to the sustainable investment, to avoid ordering effects. We checked for order effects, but we did not find any; see Table A.1.

for the sustainable investment. This procedure allows us to measure WTP values ranging from €−78.125 to €+78.125 with a precision of €1.25.

A.1.2 Incentivization. After we have elicited each participant's stated WTP for his or her preferred investment, we draw a random amount between the highest and lowest WTP that we can detect with our design—that is, between €−78.125 and €+78.125. If this random amount is smaller than the elicited WTP, we invest €1,000 in the investor's chosen investment, deducting the randomly drawn amount. If the random amount is larger than or equal to the elicited WTP, we invest €1,000 in the other investment, as is standard in the BDM mechanism. This is to ensure that investors have no incentive to deviate from their true preferences (Becker, Degroot, and Marschak 1964).

A.1.3 Impact realization. We realize the impact component of the selected investment by purchasing the corresponding amount of carbon credits from a verified greenhouse gas emission reduction project. Such projects save greenhouse gas emissions, which are measured in tons of CO_2 equivalents, according to standardized methodologies (e.g., as defined by the Verified Carbon Standard or the Gold Standard). All emissions savings are verified by an independent third party. As it has been questioned in various cases whether investments in emission reduction projects lead to the claimed emissions savings (Alexeew et al. 2010), we implement our emissions savings with a project that avoids methane emissions from organic waste treatment in Vietnam.[27] In a report commissioned by the European Commission, Cames et al. (2016) conclude that, as opposed to other project types, methane reduction projects are highly likely to lead to the claimed emissions reductions.

A.2 Detailed procedure: MTurk replications

The MTurk replications largely follow the original experimental procedure we have used in our main experiment. All deviations from the original procedure are reported for each replication in the corresponding subsection below. To ensure comparability with our main results, all replications are incentivized in the same way: We randomly select 10 participants among all MTurk participants, and make a real $1,000 investment for each of them, following the procedure described under A.1.2.

A.2.1 R1: MTurk Baseline. In this experiment, we replicate our main experiment with a sample of MTurks. We follow the original experimental design, adjusting the currency from EUR to USD. We recruit MTurk participants over the Approved Participants Panel offered by Cloud Research. MTurks in this panel have delivered evidence of a high level of engagement and attention. In addition, we restrict our study to MTurks living in the United States with a substantial track record of completing tasks well (at least 1,000 completed tasks on the platform and an acceptance rate for these tasks of at least 96%). As a further measure to ensure data quality, we add an attention check early in the experiment. As in the initial experiment, we exclude participants who do not agree that climate change is a serious problem that needs to be solved (stating an agreement of 3 or less on a scale of 1–7). We also exclude participants who explicitly disagree with the elicited WTP and are unwilling to repeat the elicitation. In addition, we apply the following exclusion criteria: We exclude participants who fail the attention check; we exclude participants who fail more than once in at least one of the two quizzes; and we exclude participants who take less than four minutes to complete the experiment. The same recruitment and quality control procedures are applied for all MTurk replications described below. For this experiment, we collect observations until we have exactly 1,000 observations that pass the above-mentioned screening criteria. The preregistration of this replication can be found under: https://aspredicted.org/M34_P2R.

[27] For more detailed information on the project, see https://market.southpole.com/home/offset-emissions/project-details/71.

A.2.2 R2: Randomized Risk & Return. In this replication, we introduce two randomized treatment conditions regarding the past risk and return of the sustainable investment to the experimental design used in our baseline MTurk experiment (R1). In both risk/return treatment groups, the conventional investment features an "average" Morningstar Risk rating and an annualized past return of 6%, as in the original experimental design. In the first treatment condition regarding risk and return, the sustainable investment features a "below-average" Morningstar Risk rating and an annualized past return of 5%. The second fund features an "above-average" Morningstar Risk rating and an annualized past return of 7%. In addition to the original design, we provide the information that a lower Morningstar Risk rating corresponds to a more favorable risk profile to avoid confusion about the directionality of this measure. Also, to find real mutual funds matching these characteristics, we broaden the fund category from "US Large-Cap Blend Equity" to "US Large-Cap Equity". We collect observations until we have precisely 400 observations that pass the screening criteria described above under "R1: MTurk Baseline". The preregistration of this replication can be found under: https://aspredicted.org/B2P_PTF.

A.2.3 R3: Direct Ask (No Anchor). In this replication, instead of applying the discrete choice method of the original design, we directly ask investors how much more they are willing to pay for the sustainable investment. Specifically, we ask the following question: "Please consider a $1,000 investment in one of the two funds presented above. How much more would you be willing to pay for Fund A (B) compared to Fund B (A) in terms of upfront fees? Please enter an amount in $:," where Fund A (B) corresponds to the sustainable investment. Participants receive the identical fund information as in the baseline MTurk experiment (R1). We collect observations until we have precisely 200 observations that pass the above screening criteria under "R1: MTurk Baseline." The preregistration of this replication can be found under: https://aspredicted.org/JXQ_ZGD.

A.2.4 R4: Direct Ask (Low Anchor). In this replication, we follow the experimental setup described under "R3: Direct Ask (No Anchor)", however providing information on common fees for equity funds. Specifically, we display the following statement directly before the question on investors' WTP: "A standard passive US Large Cap Equity fund such as Fund A (B) charges a fee of about 0.1% per year. This is $1 for a $1,000 investment.", where Fund A (B) corresponds to the conventional investment. We collect observations until we have exactly 200 that pass the screening criteria described above under "R1: MTurk Baseline." The preregistration of this replication can be found under: https://aspredicted.org/JXQ_ZGD.

A.2.5 R5: Direct Ask (High Anchor). In this replication, we follow the experimental setup described under "R3: Direct Ask (No Anchor)," however, providing information on common fees for equity funds. Specifically, we display the following statement directly before the question on investors' WTP: "A standard active US Large Cap Equity fund such as Fund A (B) charges a fee of about 1% per year. This is $10 for a $1,000 investment," where Fund A (B) corresponds to the conventional investment. We collect observations until we have exactly 200 observations that pass the screening criteria described above under "R1: MTurk Baseline." The preregistration of this replication can be found under: https://aspredicted.org/JXQ_ZGD.

A.2.6 R6: No Neutral Choice. In this replication, we follow the experimental setup described under "R1: MTurk Baseline." However, we omit the first investment choice, in which participants decide between the two funds with upfront fees that are equal. In the first choice, the sustainable fund already costs $40 more in terms of upfront fees than the conventional one. We collect observations until we have exactly 200 that pass the screening criteria described above under "R1: MTurk Baseline." The preregistration of this replication can be found under: https://aspredicted.org/GZL_LCN.

A.2.7 R7: Upward Elicitation. In this replication, instead of applying the midpoint method of our main experiment, we use a discrete choice method that gradually increases the fee differences between the sustainable investment and the conventional one. As in the main experiment, participants first face a choice between the two investments in which, for both investments, an upfront fee of $10 is deducted. In the next choice, we increase the upfront fee for the investment preferred in the first choice by $1.25. If a participant still prefers the same investment with this level of fees, we increase the upfront fee of this investment so that the difference in fees between the two investments doubles. So, in the third choice, the fee for the preferred investment would be $2.5 higher than the one for the other investment. As long as a participant keeps selecting the initially preferred investment, we keep doubling the fee differences. Once a participant switches to the other investment in any of his choices, we start applying the bisection method used in the original setup until we reach a precision of $1.25 for the WTP elicitation. If the difference in fees exceeds $80, we directly ask participants for the maximum differences in fees they would be willing to pay for the preferred fund, analogous to the procedure of the main experiment. We collect observations until we have exactly 200 that pass the screening criteria described above under "R1: MTurk Baseline." The preregistration of this replication can be found under: https://aspredicted.org/4KB_V7F.

A.2.8 R8: Lower Scale. In this replication, we apply the midpoint method used in our main experiment, however, using a lower measurement scale. After the first choice, in which participants face a choice between the two investments with an upfront fee of $10 for both investments, we increase the fee of the preferred investment by $20. This is opposed to the main experiment, in which we increase the fee for the preferred investment by $40 in the second choice. As in the original procedure, investors make seven investment choices. For participants that do not deviate from the initially preferred fund in all subsequent choices, we directly ask for the maximum differences in fees they would be willing to pay for the preferred fund, as we do in the main experiment. This setup increases our measurement precision for investors' WTP by a factor of two and reduces the maximal range of our discrete choices by a factor of two. We collect observations until we have exactly 200 that pass the screening criteria described above under "R1: MTurk Baseline." The preregistration of this replication can be found under: https://aspredicted.org/GZL_LCN.

A.2.9 R9: Higher Scale. In this replication, we apply the midpoint method used in our main experiment, however, using an increased measurement scale. After the first choice, in which participants face a choice between the two investments with an upfront fee of $10 for both investments, we increase the fee of the preferred investment by $60. This is opposed to the main experiment, in which we increase the fee for the preferred investment by $40 in the second choice. As in the original procedure, investors make seven investment choices. For investors who do not deviate from the initially preferred fund in all subsequent choices, we directly ask for the maximum differences in fees they would be willing to pay for the preferred fund, as we do in the main experiment. This setup decreases our measurement precision for investors' WTP by a factor of two, but also increases the maximal range of our discrete choices by a factor of two. We collect observations until we have exactly 200 that pass the screening criteria described above under "R1: MTurk Baseline.". The preregistration of this replication can be found under: https://aspredicted.org/PNQ_L29.

Table A.1
Order effects

	Private Investors		MTurks	
	(1) WTP	(2) WTP	(3) WTP	(4) WTP
Right-hand design	−3.892	−9.362	0.0539	−0.146
	(5.680)	(8.090)	(2.135)	(2.969)
Impact treatment	6.838	1.514	1.952	1.748
	(5.680)	(7.981)	(2.136)	(2.996)
Right-hand design * Impact treatment		10.79		0.415
		(11.36)		(4.275)
Constant	44.14***	46.45***	42.95***	43.04***
	(4.661)	(5.260)	(1.779)	(2.019)
Observations	196	196	1,000	1,000
R^2	0.009	0.013	0.001	0.001
F	0.859	0.873	0.420	0.283

* $p < .05$; ** $p < .01$; *** $p < .001$.
This table presents the results of an ordinary least squares (OLS) regression with WTP for the sustainable investment as the dependent variable. In both specifications, an indicator variable taking the value of 0 for participants seeing the sustainable investment on the left-hand side and 1 for participants seeing the sustainable investment on the right-hand side is included, together with an indicator variable taking the value of 0 for the LowImpact treatment and 1 for the HighImpact treatment. Specification (1) reports the results for our sample of private investors in the main experiment; specification (2) reports the results for the baseline experiment with MTurks ("R1: MTurk Baseline"). Standard errors are shown in parentheses.

Table A.2
Main results, excluding investors who fail the comprehension quizzes

	Mean Values		Mann–Whitney U Test
	LowImpact (n=91)	HighImpact (n=94)	(HighImpact = LowImpact)
WTP (€)	42.66	48.77	p=.383
WTP/Impact (€/tCO_2)	85.32	9.75	p=.001

This table presents private investors' absolute and relative WTP for the sustainable fund, as elicited in the main experiment, but excluding investors who twice fail at least one of the two quizzes that test their comprehension of the information that is provided to them. The first two columns report mean values of the variables, by impact treatment; the third column reports p-values of a Mann–Whitney U test, testing for differences between the two treatments.

Table A.3
Main results excluding investors with censored WTP

	Mean Values		Mann–Whitney U Test
	LowImpact (n=81)	HighImpact (n=74)	(HighImpact = LowImpact)
WTP (€)	30.24	28.22	p=.903
WTP/Impact (€/tCO_2)	60.48	5.64	p=.001

This table presents private investors' absolute and relative WTP for the sustainable investment, as elicited in the main experiment, excluding investors with censored WTP. Investors' WTP is censored if they do not deviate from the initially preferred investment in all seven investment choices. We cannot elicit WTP directly from the choices of these investors. We additionally ask these investors to state their WTP for the sustainable investment. These stated values are excluded from the results presented in this table. For each treatment group, the mean WTP is lower when excluding investors with censored WTP. The first two columns report mean values of the variables, by impact treatment; the third column reports p-values of a Mann–Whitney U test, testing for differences between the two treatments.

Table A.4
Definition of variables elicited in the postexperiment survey

Variable	Measure
Risk expectations	Answer to the question "How do you expect that Fund A and Fund B compare in terms of risk?" on a 5-point scale from "Investing in Fund A is much riskier" to "Investing in Fund B is much riskier." Values are normalized to a scale from −10 to 10, where positive values indicate that investors expect the sustainable investment to be less risky, and negative ones that they expect the conventional investment to be less risky.
Return expectations	Answer to the question "How do you expect that Fund A and Fund B compare in terms of return?" on a 5-point scale from "Fund A will deliver a much higher return" to "Fund B will deliver a much higher return." Values are normalized to a scale from −10 to 10, where positive values indicate that investors expect the sustainable investment to deliver higher returns, and negative ones that they expect the conventional investment to deliver higher returns.
Positive emotions	Answer to the question "How do Fund A and Fund B compare in terms of how it feels to invest in the fund?" on a 5-point scale from "It feels much better to invest in Fund A" to "It feels much better to invest in Fund B." Values are normalized to a scale from −10 to 10, where positive values indicate that it feels better for investors to choose the sustainable investment, and negative values that it feels better to choose the conventional investment.
Perceived investment impact	Agreement with the statement "Investing in Fund [A,B] makes a relevant contribution to fighting climate change," on a 7-point Likert scale. Values are normalized to a scale from −10 to 10.
General relevance impact	Agreement with the statement "When investing, it is important to me whether I contribute to fighting climate change" on a 7-point Likert scale. Values are normalized to a scale from −10 to 10.
General relevance impact level	Agreement with the statement "When investing, it is important to me *how much* I contribute to fighting climate change" on a 7-point Likert scale, where [A,B] corresponds to the sustainable investment. Values are normalized to a scale from −10 to 10.
Estimated cost of saving 1 ton of CO_2	Answer to the question "What do you think: What are the average costs of saving 1,000 kg of CO_2 emissions (in €)?" To reduce the influence of extreme values, values are winsorized at the 5% and 95% levels.
Risk preferences	Answer to the question "In general, how willing or unwilling are you to take risks?" on a 10-point scale (1 = "Completely unwilling to take risks"; 10 = "Very willing to take risks"), according to the experimentally validated survey module of Falk et al. (2016).
Time preferences	Answer to the question "How willing are you to give up something that is beneficial for you today in order to benefit more from that thing in the future?" on a 10-point scale (1 = "Completely unwilling"; 10 = "Very willing to do so"), following Falk et al. (2016).
Altruism	Answer to the question "How do you assess your willingness to share with others without expecting anything in return?" on a 10-point scale (1 = "Completely unwilling to share"; 10 = "Very willing to share"), following Falk et al. (2016).
Climate awareness	Agreement with the statement "Climate change is a serious problem that needs to be solved," on a 7-point Likert scale. Values are normalized to a scale from −10 to 10.
Female	The dummy variable *Female* takes the value of 1 if the investor chooses *Female* from among the options *Female*, *Male*, and *Other*, and 0 if not.
Age	Investor's self-stated age.
Income	Self-reported annual household income, with options ranging from "less than €10,000" to "€200,000 or more" in steps of €5,000.
Net worth	Self-reported household net worth with seven options ranging from "less than €50,000" to "more than €10 million."
Highest education	Self-reported highest degree or level of schooling the investor has completed.
Investment knowledge	Agreement with the statement "Compared to the average of the population, my investment knowledge is good" on a 7-point Likert scale (1 = "Strongly disagree"; 7 = "Strongly agree"), following Dorn and Huberman (2005) and Riedl and Smeets (2017). Values are normalized to a scale from 0 to 10.

This table shows the definitions and measurement of the variables we elicit in our postexperiment survey.

Table A.5
Risk and return expectations and investors' WTP for sustainability

	Private Investors		Impact Investors	
	(1) WTP	(2) WTP	(3) WTP	(4) WTP
Risk expectations	0.252	0.0689	−1.386	−2.258
	(0.891)	(0.873)	(1.288)	(1.288)
Return expectations	0.851	0.745	−0.712	−0.126
	(0.710)	(0.695)	(1.116)	(1.188)
Impact treatment	7.450	6.739	1.204	1.970
	(5.658)	(5.567)	(6.020)	(6.107)
Demographics	No	Yes	No	Yes
Constant	41.94***	9.232	49.44***	−34.23
	(4.060)	(31.38)	(4.335)	(31.98)
Observations	194	194	118	118
R^2	0.017	0.120	0.020	0.121
F	1.078	2.058	0.779	1.204

* $p < .05$; ** $p < .01$; *** $p < .001$.
This table presents the results of an ordinary least squares (OLS) regression with the WTP for the sustainable investment as the dependent variable. In all specifications, investors' risk and return expectations for the sustainable investment, as elicited in our postexperiment survey, are included as independent variables. Both risk and return expectations are transformed to a scale from −10 to 10, where positive values indicate that investors have a more favorable view of the sustainable investment, negative ones that they have a more favorable view of the conventional investment. All specifications include an indicator variable taking the value of 0 for the LowImpact treatment and 1 for the HighImpact treatment. In addition, specifications (2) and (4) also include controls for investors' demographics, as described in detail in Table A.4. Specifications (1) and (2) report the results for our sample of private investors in the main experiment; specifications (3) and (4) report the results for our sample of impact investors. Standard errors are shown in parentheses.

Table A.6
WTP before and after the onset of the COVID-19 crisis

	Mean Values		Mann–Whitney U Test
	LowImpact	HighImpact	(HighImpact = LowImpact)
September 2019			
N	159	152	
WTP (€)	27.64	29.82	$p = .533$
WTP/Impact (€/tCO_2)	55.28	5.96	$p < .001$
September 2020			
N	119	123	
WTP (€)	32.03	27.85	$p = .262$
WTP/Impact (€/tCO_2)	64.04	5.57	$p < .001$

This table reports the results of a preliminary version of our experiment, which we ran with students at Radboud University in September 2019, well before the emergence of SARS-CoV-2, as well as in September 2020. The first two columns report mean values of the variables, by impact treatment; the third column reports p-values of a Mann–Whitney U test, testing for differences between the two treatments.

Table A.7
Results for the impact investors during the preregistration period

	Mean Values		Mann–Whitney *U* Test
	LowImpact ($n = 56$)	HighImpact ($n = 57$)	(HighImpact = LowImpact)
WTP (€)	48.79	49.45	$p = .850$
WTP/Impact (€/tCO_2)	97.59	9.89	$p < .001$

This table presents investors' absolute and relative WTP for the sustainable investment, as elicited in our experiment with the sample of impact investors, strictly following the preregistered procedure and thereby excluding five investors who participated in the experiment starting more than 3 months after its launch. The first two columns report mean values of the variables, by impact treatment; the third column reports *p*-values of a Mann–Whitney *U* test, testing for differences between the two treatments.

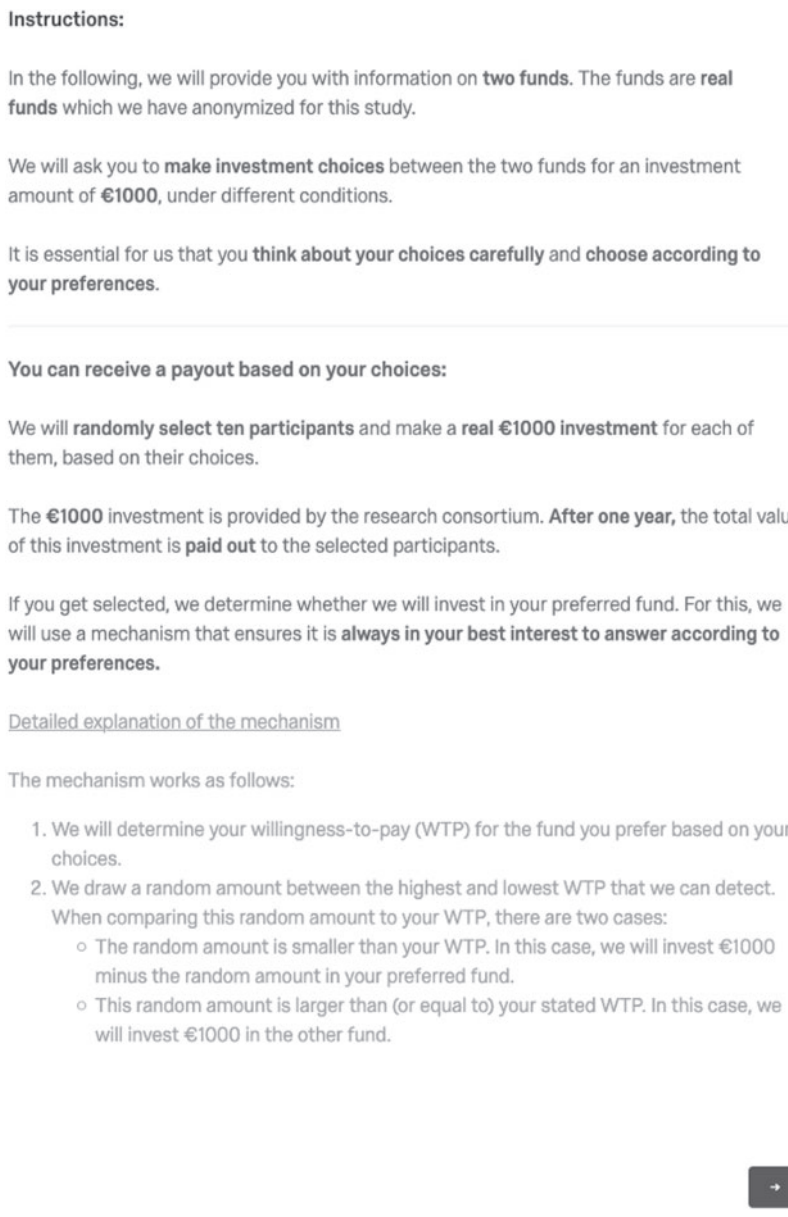


Figure A.1
Instructions
This figure shows a screenshot of the instructions that investors receive.

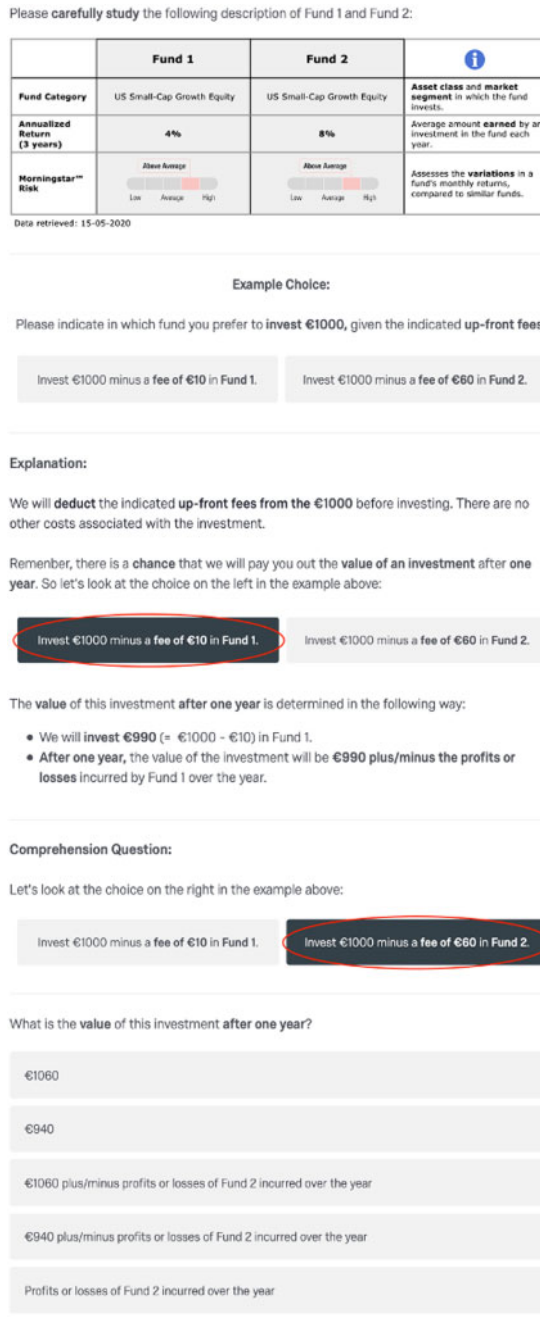
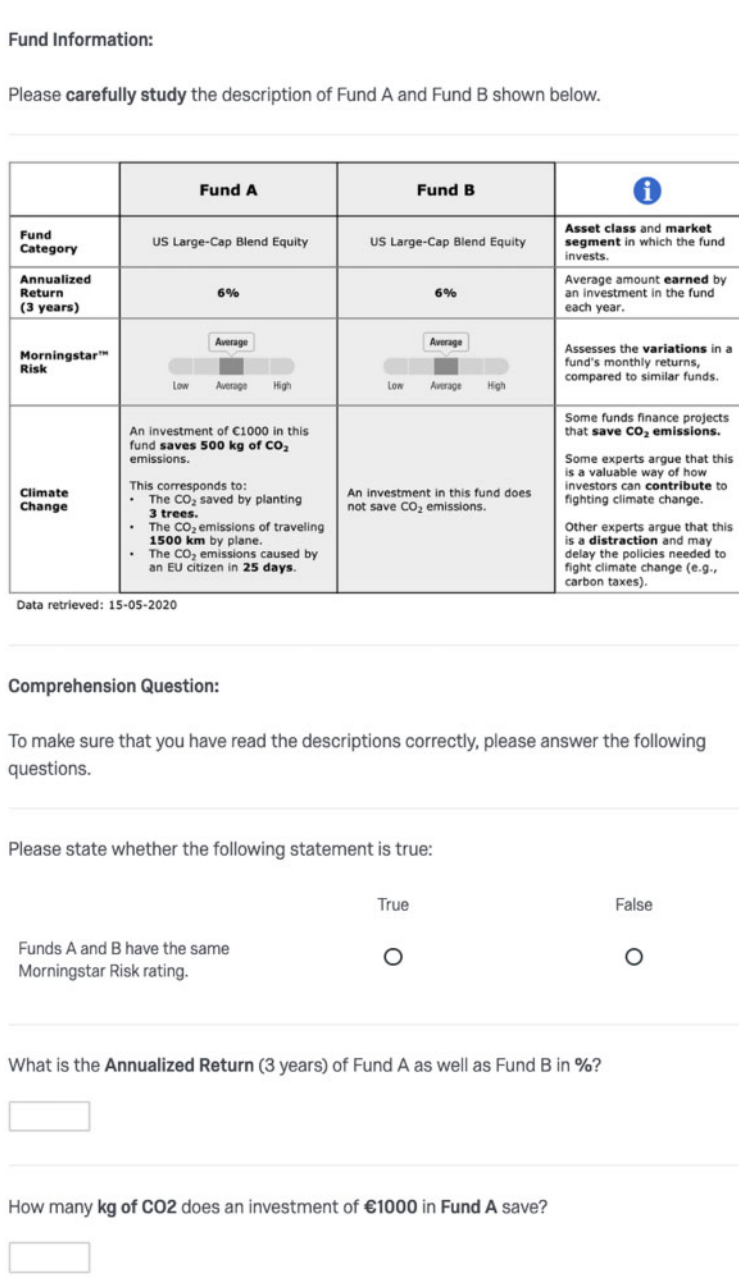


Figure A.2
Example choice
This figure shows a screenshot of the quiz investors participate in before making the investment decisions.



Fund Information:

Please **carefully study** the description of Fund A and Fund B shown below.

	Fund A	Fund B	ℹ
Fund Category	US Large-Cap Blend Equity	US Large-Cap Blend Equity	**Asset class** and **market segment** in which the fund invests.
Annualized Return (3 years)	**6%**	**6%**	Average amount **earned** by an investment in the fund each year.
Morningstar™ Risk	Average Low Average High	Average Low Average High	Assesses the **variations** in a fund's monthly returns, compared to similar funds.
Climate Change	An investment of €1000 in this fund **saves 500 kg of CO_2** emissions. This corresponds to: • The CO_2 saved by planting **3 trees.** • The CO_2 emissions of traveling **1500 km** by plane. • The CO_2 emissions caused by an EU citizen in **25 days**.	An investment in this fund does not save CO_2 emissions.	Some funds finance projects that **save CO_2 emissions.** Some experts argue that this is a valuable way of how investors can **contribute** to fighting climate change. Other experts argue that this is a **distraction** and may delay the policies needed to fight climate change (e.g., carbon taxes).

Data retrieved: 15-05-2020

Comprehension Question:

To make sure that you have read the descriptions correctly, please answer the following questions.

Please state whether the following statement is true:

	True	False
Funds A and B have the same Morningstar Risk rating.	○	○

What is the **Annualized Return** (3 years) of Fund A as well as Fund B in **%**?

How many **kg of CO2** does an investment of **€1000** in **Fund A** save?

Figure A.3
Investment information and comprehension quiz
This figure provides a screenshot of the information the investors participating in our main experiment receive on the two investments if they are assigned to the LowImpact treatment, as well as of the comprehension quiz investors have to "win" if they are to continue.

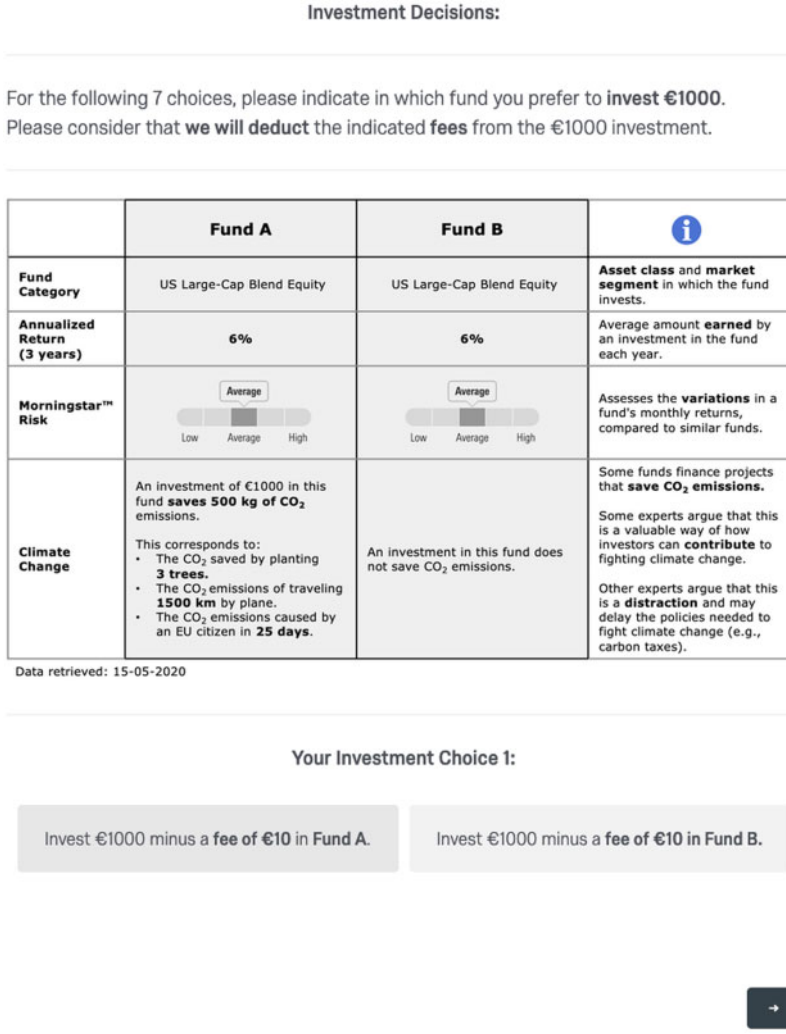


Figure A.4
Investment choice
The screenshot in this figure shows an example of the investment choices the investors face.

	Fund A	Fund B	Fund C	
Fund Category	US Large-Cap Blend Equity	US Large-Cap Blend Equity	US Large-Cap Blend Equity	**Asset class** and **market segment** in which the fund invests.
Annualized Return (3 years)	**6%**	**6%**	**6%**	Average amount **earned** by an investment in the fund each year.
Morningstar™ Risk	Average Low Average High	Average Low Average High	Average Low Average High	Assesses the **variations** in a fund's monthly returns, compared to similar funds.
Climate Change	An investment into Fund A does not save CO_2 emissions.	An investment of €1000 in this fund **saves 50 kg of CO_2** emissions. This corresponds to: • **30%** of the CO_2 saved by planting a **tree.** • The CO_2 emissions of traveling **150 km** by plane. • The CO_2 emissions caused by an EU citizen in **2.5 days.**	An investment of €1000 in this fund **saves 500 kg of CO_2** emissions. This corresponds to: • The CO_2 saved by planting **3 trees.** • The CO_2 emissions of traveling **1500 km** by plane. • The CO_2 emissions caused by an EU citizen in **25 days.**	Some funds finance projects that **save CO_2 emissions.** Some experts argue that this is a valuable way of how investors can **contribute** to fighting climate change. Other experts argue that this is a **distraction** and may delay the policies needed to fight climate change (e.g., carbon taxes).

Data retrieved: 15-05-2020

Figure A.5
Screenshot of the investment information in the joint evaluation extension
This figure provides an example of the information the investors receive in the joint evaluation extension of our experiment on the three investment options. The screenshot corresponds to the investment information investors in the LowImpactRange treatment receive.

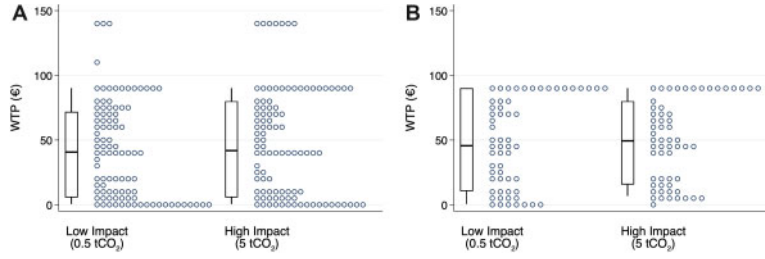


Figure A.6
Distribution of investors' WTP for sustainable investments
This figure shows vertical histograms of investors' WTP for the sustainable investment. Panel A shows the WTP in the main experiment with private investors, Panel B shows the WTP of impact investors. Each dot reflects one individual WTP, and dots are grouped and stacked to the right in brackets with a width of €5. The boxes on the left show the 10, 25, 50, 75, and 90 percentiles. Note that the data has been winsorized at the 5% and 95% levels, according to the preregistration.

References

Abdellaoui, M. 2000. Parameter-free elicitation of utility and probability weighting functions. *Management Science* 46:1497–512.

Aguinis, H., I. Villamor, and R. S. Ramani. 2021. MTurk research: Review and recommendations. *Journal of Management* 47:823–37.

Alexeew, J., L. Bergset, K. Meyer, J. Petersen, L. Schneider, and C. Unger. 2010. An analysis of the relationship between the additionality of CDM projects and their contribution to sustainable development. *International Environmental Agreements: Politics, Law and Economics* 10:233–48.

Andreoni, J. 1989. Giving with impure altruism: Applications to charity and Ricardian equivalence. *Journal of Political Economy* 97:1447–58.

———. 1990. Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal* 100:464–77.

Andreoni, J., and J. Miller. 2002. Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica* 70:737–53.

Arechar, A. A., S. Gächter, and L. Molleman. 2018. Conducting interactive experiments online. *Experimental Economics* 21:99–131.

Barber, B. M., A. Morse, and A. Yasuda. 2021. Impact investing. *Journal of Financial Economics* 139:162–85.

Bauer, R., T. Ruof, and P. Smeets. 2021. Get real! Individuals prefer more sustainable investments. *Review of Financial Studies* 34:3976–4043.

Becker, G. M., M. H. Degroot, and J. Marschak. 1964. Measuring utility by a single-response sequential method. *Behavioral Science* 9:226–32.

Becker, G. S. 1974. A theory of social interactions. *Journal of Political Economy* 82:1063–93.

Benabou, R., and J. Tirole. 2010. Individual and corporate social responsibility. *Economica* 77:1–19.

Bonnefon, J.-F., A. Landier, P. Sastry, and D. Thesmar. 2022. The moral preferences of investors: Experimental evidence. Working Paper, National Bureau of Economic Research.

Broccardo, E., O. Hart, and L. Zingales. 2020. Exit vs. voice. Working Paper, National Bureau of Economic Research.

Brodback, D., N. Günster, and S. Pouget. 2021. The valuation of corporate social responsibility: A willingness-to-pay experiment. Working Paper, Network for Studies on Pensions, Aging and Retirement.

Busch, T., P. Bruce-Clark, J. Derwall, R. Eccles, T. Hebb, A. Hoepner, C. Klein, P. Krueger, F. Paetzold, B. Scholtens, and O. Weber. 2021. Impact investments: A call for (re)orientation. *SN Business & Economics* 1:1–33.

Cames, M., R. O. Harthan, J. Füssler, M. Lazarus, C. M. Lee, P. Erickson, and R. Spalding-Fecher. 2016. How additional is the clean development mechanism? Working Paper, Öko-Institut.

Coval, J. D., and T. J. Moskowitz. 1999. Home bias at home: Local equity preference in domestic portfolios. *Journal of Finance* 54:2045–73.

Desvousges, W. H., F. R. Johnson, R. W. Dunford, K. J. Boyle, S. P. Hudson, and K. N. Wilson. 1992. *Measuring nonuse damages using contingent valuation: An experimental evaluation of accuracy*. Durham: RTI Press.

Dorn, D., and G. Huberman. 2005. Talk and action: What individual investors say and what they do. *Review of Finance* 9:437–81.

Eckel, C. C., and P. J. Grossman. 1996. Altruism in anonymous dictator games. *Games and Economic Behavior* 16:181–91.

Falk, A., A. Becker, T. J. Dohmen, D. Huffman, and U. Sunde. 2016. The preference survey module: A validated instrument for measuring risk, time, and social preferences. Working Paper, IZA.

Fama, E. F., and K. R. French. 2007. Disagreement, tastes, and asset prices. *Journal of Financial Economics* 83:667–89.

Ferguson, E., and N. Flynn. 2016. Moral relativism as a disconnect between behavioural and experienced warm glow. *Journal of Economic Psychology* 56:163–75.

Finucane, M. L., A. Alhakami, P. Slovic, and S. M. Johnson. 2000. The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making* 13:1–17.

Gibson, R., S. Glossner, P. Krueger, P. Matos, and T. Steffen. Forthcoming. Do responsible investors invest responsibly? *Review of Finance*.

Global Sustainable Investment Alliance. 2021. Global sustainable investment review 2020. http://www.gsi-alliance.org/wp-content/uploads/2021/08/GSIR-20201.pdf.

Harrison, G. W., and J. A. List. 2004. Field experiments. *Journal of Economic Literature* 42:1009–55.

Hartzmark, S. M., and A. B. Sussman. 2019. Do investors value sustainability? A natural experiment examining ranking and fund flows. *Journal of Finance* 74:2789–837.

Heinkel, R., A. Kraus, and J. Zechner. 2001. The effect of green investment on corporate behavior. *Journal of Financial and Quantitative Analysis* 36:431–49.

Holt, C. A., and S. K. Laury. 2014. Assessment and estimation of risk preferences. In M. Machina and K. Viscusi, eds., *Handbook of the Economics of Risk and Uncertainty*, 135–201. Amsterdam: Elsevier.

Hong, H., N. Wang, and J. Yang. 2021. Welfare consequences of sustainable finance. Working Paper, National Bureau of Economic Research.

Hsee, C. K., and Y. Rottenstreich. 2004. Music, pandas, and muggers: On the affective psychology of value. *Journal of Experimental Psychology: General* 133:23–30.

Hsee, C. K., and J. Zhang. 2010. General evaluability theory. *Perspectives on Psychological Science* 5:343–55.

Huberman, G. 2001. Familiarity breeds investment. *Review of Financial Studies* 14:659–80.

Humphrey, J., S. Kogan, J. Sagi, and L. T. Starks. 2020. The asymmetry in responsible investing preferences. Working Paper, National Bureau of Economic Research.

Ivković, Z., and S. Weisbenner. 2009. Individual investor mutual fund flows. *Journal of Financial Economics* 92:223–37.

Kahneman, D., and J. L. Knetsch. 1992. Valuing public goods: The purchase of moral satisfaction. *Journal of Environmental Economics and Management* 22:57–70.

Karlan, D., and D. H. Wood. 2017. The effect of effectiveness: Donor response to aid effectiveness in a direct mail fundraising experiment. *Journal of Behavioral and Experimental Economics* 66:1–8.

Khorana, A., H. Servaes, and P. Tufano. 2009. Mutual fund fees around the world. *Review of Financial Studies* 22:1279–310.

Kuhnen, C. M., and B. Knutson. 2011. The influence of affect on beliefs, preferences, and financial decisions. *Journal of Financial and Quantitative Analysis* 46:605–26.

Landier, A., and S. Lovo. 2020. ESG investing: How to optimize impact? Working Paper, HEC Paris.

Laudi, M., P. Smeets, and U. Weitzel. 2021. Do financial advisors exploit responsible investment preferences? Working Paper, Maastricht University.

MacGregor, D. G., P. Slovic, D. Dreman, and M. Berry. 2000. Imagery, affect, and financial judgment. *Journal of Psychology and Financial Markets* 1:104–10.

Metzger, L., and I. Günther. 2019. Making an impact? The relevance of information on aid effectiveness for charitable giving. A laboratory experiment. *Journal of Development Economics* 136:18–33.

Null, C. 2011. Warm glow, information, and inefficient charitable giving. *Journal of Public Economics* 95:455–65.

Oehmke, M., and M. M. Opp. 2019. A theory of socially responsible investment. Working Paper, Swedish House of Finance.

Pastor, L., R. F. Stambaugh, and L. A. Taylor. 2021. Sustainable investing in equilibrium. *Journal of Financial Economics* 142:550–71.

Pedersen, L. H., S. Fitzgibbons, and L. Pomorski. 2021. Responsible investing: The ESG-efficient frontier. *Journal of Financial Economics* 142:572–97.

Pindyck, R. S. 2019. The social cost of carbon revisited. *Journal of Environmental Economics and Management* 94:140–60.

Ramelli, S., and A. F. Wagner. 2020. Feverish stock price reactions to COVID-19. *Review of Corporate Finance Studies* 9:622–55.

Riedl, A., and P. Smeets. 2017. Why do investors hold socially responsible mutual funds? *Journal of Finance* 72:2505–50.

Slovic, P., M. L. Finucane, E. Peters, and D. G. MacGregor. 2007. The affect heuristic. *European Journal of Operational Research* 177:1333–52.

Strong, N., and X. Xu. 2003. Understanding the equity home bias: Evidence from survey data. *Review of Economics and Statistics* 85:307–12.

United Nations Environment Programme. 2020. The emissions gap report 2020. https://www.unep.org/emissions-gap-report-2020.

van de Kuilen, G., and P. P. Wakker. 2011. The midweight method to measure attitudes toward risk and ambiguity. *Management Science* 57:582–98.

Zizzo, D. J. 2010. Experimenter demand effects in economic experiments. *Experimental Economics* 13:75–98.