The Asymmetry in Responsible Investing Preferences and Beliefs

Jacquelyn Humphrey (Business School, University of Queensland)

Shimon Kogan (Arison School of Business, Reichman University and The Wharton School)

Jacob Sagi (Kenan-Flagler Business School, University of North Carolina, Chapel Hill)

Laura Starks (McCombs School of Business, University of Texas, Austin and NBER)

September 2023

ABSTRACT: Empirical stylized facts in the literature concerning “sin” versus “angel” stocks display asymmetry. Through an experiment, we examine whether such biases can be micro-founded via individuals’ preferences and belief formations. We find that negative environmental and social externalities have thrice the impact of positive externalities on investment choices. Further, negative externalities modestly increase pessimism about investment prospects while positive externalities have no discernible impact. The asymmetry is pervasive, heterogeneous, and comparable to the magnitude observed in loss-aversion. Beyond rationalizing stylized empirical facts, our findings should help direct the growing theoretical literature that models the implications of non-pecuniary individual investor behavior.

NBER working paper 29288.

JEL Classification: G11, G41, C91

Keywords: responsible investing, sustainable investing, ESG, experiment

---

The authors thank Rob Bauer, Peter Bossaerts, Shaun Davies, Barry Oliver, Sébastien Pouget, Saphira Rekker, Paul Smeets, Gabby Walters, Paul Yoo, Josef Zechner and participants at the Inaugural Behavioral, Biological, and Experimental Economics conference, Penn State University, the Principles for Responsible Investment Academic Network conference and seminars at the Financial Research Network, University of Texas at Austin, Villanova University, Victoria University of Wellington, University of Geneva, the Wharton School and University of Woollongong for helpful comments. The authors are grateful to Terry Pan (Zheyao), Lee Seltzer, and Mitch Towner for research assistance; Larissa Garcia, Matt Pearsall, Jacob West and their team of assistants for lab assistance; and Alan McCrystal, Alex Cameron, and Paige Ottmar for programming assistance.

1. Introduction

In response to demand from their clients, institutional investors increasingly offer responsible investment (RI) products, many of which show asymmetric approaches such as boycotting firms perceived to be doing harm, but not necessarily including only firms perceived to be doing good.[1] Empirical evidence on these types of products further suggests that, in aggregate, investors have asymmetric responses to ESG information. For example, fund flows and market reactions are found to be more sensitive to negative ESG events and information than to positive (e.g., Krueger 2015; Bialkowski and Starks 2018; Hartzmark and Sussman, 2019). Similarly, in an experimental study, Chew and Li (2021) find a strong aversion to sin stocks relative to the affinity for virtue stocks. Furthermore, sin stocks trade at a significant discount, but the evidence is somewhat mixed regarding whether a premium exists for angel stocks or green bonds (Hong and Kacperczyk 2009, Larcker and Watts, 2020, Bolton and Kacperczyk, 2021, Flammer 2021). Despite the large body of empirical evidence, it is difficult to link the asymmetry to individual investors' nonpecuniary preferences and/or skewed beliefs about returns or future regulatory risks. In this paper we provide experimental evidence establishing a link between the asymmetry in responses to externalities and nonpecuniary preferences as well as return beliefs.

Our study is motivated by a large body of research demonstrating the prevalence of asymmetry in individual decision making across multiple domains. The most prominent example is the concept of loss aversion (Kahneman and Tversky 1979), in which individuals are prepared to take substantial risk to avoid the perception of loss yet are, at the same time, far more conservative when their choices are framed as gains. Other examples include the disparity between the valuation of an owned object and the price at which one is willing to

---

[1] Responsible investment (RI) is also termed environmental, social and governance (ESG) or sustainable and responsible investing (SRI). The increased interest in RI products is thought by some to be driven in part by societal expectations arising from individuals' tastes, as contrasted with their financial motivations.

acquire it (Kahneman, Knetsch, and Thaler, 1990); the more pronounced vividness of negative versus positive memories (Kensinger and Schachter, 2008); and the greater influence of anticipated regret/disappointment versus rejoicing/elation in emotion-based choice (Mellers, Schwartz, and Ritov, 1999). Correspondingly, there exists a complementary literature documenting asymmetry in how individuals process information, which also provides examples of individuals' preferences influencing their belief formation.[2] In the well-known "confirmation bias", for instance, subjects place more weight on new information that confirms their beliefs – especially when those beliefs are favorable to their worldview. More pertinent to our own contribution and mirroring preference asymmetry in loss-aversion, Kuhnen (2015) demonstrates that individuals update beliefs in an overly-pessimistic manner when payoffs are framed as losses as opposed to gains.

Although evidence exists regarding individuals' RI choice behaviors and belief formations based on non-pecuniary preferences, the evidence does not directly concern asymmetry.[3] This leaves open the question of whether asymmetry in individual RI preferences can provide a basis for observed aggregate asymmetries. Motivated by the lack of an existing link and the rich history of asymmetry in individual behavior found in other domains, we test for the existence of significant asymmetries in individual RI preferences and beliefs. Doing so is important for at least two reasons. First, it can serve to rationalize the disparate findings of RI-related aggregate asymmetries. Second, it can help to differentiate the alignment of actual investor behavior with existing (and future) models of RI preferences in theoretical studies.[4]

---

[2] See Section 9 in the review article by Benjamin (2019).

[3] For evidence, see, for example, Riedl and Smeets (2017), Hartzmark and Sussman (2019), Barber, Morse and Yasuda (2021), Bauer, Ruof, and Smeets (2021), Geczy, Jeffers, Musto and Tucker (2021), and Yoo (2022).

[4] Among the theoretical RI models, Heinkel, Kraus and Zechner (2001), Luo and Balvers (2017), Pedersen, Fitzgibbons and Pomorski (2021) and Zerbib (2022) allow for asymmetric RI preferences (e.g., negative screening). Other prominent studies do not (e.g., Chowdhry, Davies and Waters 2019; Oehmke and Opp 2020; Pastor, Stambaugh and Taylor 2021; Goldstein, Kopytov, Shen, and Xiang 2021). None consider the influence of RI preferences on beliefs.

Operationally, an asymmetric response is a disproportionate response to negative versus positive stimuli of the same magnitude. In the field, it is challenging to find settings in which "RI stimuli" can be isolated and controlled, relative to, for example the effect the stimuli can have on expectations. It is even more challenging to find a setting in which the *magnitude* of the two stimuli is identical. Thus, we study the question experimentally, which allows us to control for investors' information sets as we test the relationship between social preferences and investment decisions. Each participant's investment decisions impose externalities on nonprofit organizations ("nonprofits"), which the participants know can occur. Key to the design is our ability to control for the magnitude of the externalities while switching their sign. We separately examine whether and how RI concerns asymmetrically distort individuals' (i) preferences (holding their beliefs constant), and (ii) their beliefs, which are formed by learning from the investment outcomes.[5]

Although we find evidence supporting asymmetric influences through both channels, by far the most important is the impact of negative social outcomes on preferences. For the same dollar magnitude of externality, we estimate that subjects reduce allocations in the presence of negative externalities more than three times the amount they would increase allocations when the externality is positive. This result is reminiscent in direction and magnitude of outcome asymmetries in the other domains as discussed above, e.g., loss aversion and prospect theory (Kahneman and Tversky 1979).

In our experiment, which is adapted for an RI framework from Kuhnen's (2015) experiment, subjects receive an endowment to allocate between a risky stock and cash over multiple rounds. The stock's returns are binary (can double or halve in value) with the probability of the high outcome being fixed but unknown to subjects. Throughout the

[5] For the sake of brevity, here and elsewhere in the paper, when we refer to subjects' preferences it should be implicitly understood to mean choice behavior with beliefs held fixed.

experiment, subjects are incentivized to learn, report their beliefs about the stock's prospects and make their investment decisions.

To examine how RI considerations distort individuals' return beliefs and allocation choices relative to a neutral benchmark, we link payoffs from the stock allocation to payments that address social causes. In particular, we ask subjects to rank the importance of a list of nonprofits associated with "popular" environmental and social topics. To link the investment outcomes to the payments, we – the experimenters – make contributions to the preferred nonprofits in proportion to subjects' investment allocations. Importantly, payments to the nonprofits *do not* come out of the subjects' payoffs but are supplemental payments made by the experimenters. Thus, "neoclassical" investors – those who care only about their own payoffs when making investment decisions – would be insensitive to any link with a social cause.

The design incorporates three treatments based on the sign of the linkage between subjects' payoffs from the stock investment and payments to their preferred nonprofits. In the Neutral treatment, the two are unrelated (there are no payments to nonprofits). In the Positive (Negative) treatment, payoffs from the stock investment are matched by equal contributions (deductions) to the preferred cause.[6] The more the subject earns from their stock investment, everything else equal, the more the nonprofit receives in the Positive treatment, but the less the nonprofit receives in the Negative treatment.[7] In the Positive and Negative treatments, subjects weigh how an allocation to the risky asset will impact both their personal gains and the social goals with which they have chosen to be aligned.

[6] The Neutral treatment is always presented first to the subjects, prior to any discussion of social issues. The order of the Negative and Positive treatments is then randomized across subjects to control for priming effects in average results.

[7] In the Negative treatment, we deduct payments from non-profits that support popular non-pecuniary environmental and social causes. We acknowledge that it would have been ideal to also add/deduct payments to non-profits that advocate *against* these types of causes (e.g., the National Rifle Association) but institutional sensitivities around contributions to controversial non-profits and their inevitable linkage to the study made such a choice infeasible. We address this shortcoming in Section 3.3.

This design affords two important advantages. First, we are able to keep the magnitude of the externalities constant across treatments. Thus, we can compare behavior not just against the Neutral treatment but also between the Positive and Negative treatments in a way that quantitatively controls for the external stimulus. The design mimics a secondary market environment facing atomistic investors whose individual actions have insignificant impact on the firms being traded. Second, the treatments only distort investors' non-pecuniary motives and, in so doing, allow us to study how they may be incorporated into an investment context. Under neoclassical assumptions, allocations and beliefs should be similar across the different treatments. That is, the design rules out common alternative financial motives for RI preferences, such as expectations about higher returns or lower risk.

The experiment generates round-by-round data on individuals' elicited beliefs over stock payoffs as well as data on their investment allocation choices. We document a statistically strong effect from negative RI externalities – reported beliefs and allocations are significantly different from the baseline. Moreover, the influence on allocation decisions is economically pronounced (roughly a 30% reduction in average allocation across subjects), though much less so over beliefs formation – roughly a 2% reduction in the perceived probability that the stock's returns are positive, consistent with Kuhnen's (2015) findings for updating in a "loss" frame. In stark contrast, for positive RI externalities we find far less internalization on either allocation decisions or beliefs. Experimental outcomes from the Positive treatment resemble those from the Neutral treatment. Overall, the asymmetric impact of negative externalities on preferences and beliefs is similar in magnitude to asymmetries observed in studies of behavior in loss versus gain frames (e.g., loss aversion).

The weak Positive treatment effects we observe are especially surprising given that a simple charitable-giving motive is prevalent in practice and in experimental economics. Thus, while an investment context may mask or muddle charitable-giving behavior, negative RI stimulus results in a pronounced aversive response.

One potential concern with our experimental design is the validity of the mapping between the Negative treatment to choices in the field. In the Negative treatment, payoffs from investment are associated with a reduction in the contribution to a nonprofit that works towards positive externalities. One may be concerned that this design does not capture the choice investors face in practice because, for example, sin stock investments can result in increased payoffs from negative externalities. To address this potential concern, we design a subsequent survey that uses the same set of causes as in the main experiment and ask subjects to state their preferences between contributing funds towards a negative cause versus reducing the funds of a charity opposing that cause. Respondents, (of similar demographics and gender to the original experiment's population), exhibit a strong preference for reducing funding from a charity doing good rather than contributing funds that will be used to do harm. Thus, the results in our main experiment likely *understate* the degree of asymmetry in the field.

Beyond the average treatment effect, we find significant heterogeneity across subjects' RI sensitivities to negative but not positive externalities. While subjects' responses to positive social externalities appear to come from a population that is only weakly attuned to positive RI stimulus, the negative externality splits the subject population into two sub-populations: One weakly and one strongly sensitive to the negative RI stimulus. In the Negative treatment, if the accumulated information about the stock is positive, half of the subjects would invest an average of roughly one half of their Neutral treatment allocation. In an appendix, we use within-subject data and a reduced-form structural model to estimate heterogeneous allocation curves as functions of investors' reported probabilities in the three treatments. These estimates can potentially be used to construct or calibrate more descriptive theoretical models RI of behavior.[8]

In Section 2 we outline the experiment, reporting the details of the analyses and their interpretations in Section 3. In Section 4 we provide a discussion of our paper's contribution to the literature in light of prior work. Section 5 concludes.

2. Experimental design

2.1 Description of experiment

The experiment is organized around a basic set of tasks we term a "trial" performed through a computer terminal.[9] The experiment itself consists of a series of trials, some of which include treatment effects. Before describing the experiment's sequence of events, we detail the mechanics of a single trial.

At the beginning of each trial, participants are informed that a stock in which they can invest during that trial may be one of two types: a high payoff stock that doubles the amount invested with a probability of 2/3 and halves the investment with a probability of 1/3; or a low payoff stock that doubles investment with a probability of 1/3 and halves it with a probability of 2/3. Participants are also informed that the computer randomly selects the stock type at the beginning of the trial, with equal probability, and that the stock's type remains fixed for the duration of the trial. The trial consists of six rounds of investment during which subjects can learn about the stock's type. Although the investment payoff outcome is disclosed at the end of each round, the stock type is not disclosed.

[8]Most models of non-pecuniary RI investing are limited to two types of investors: Those with only financial concerns and those that additionally exhibit non-pecuniary preferences. The latter are typically archetypes who either screen stocks (extreme asymmetry) or exhibit no asymmetric RI preferences.
[9] Each subject participates in the experiment through a distinct computer terminal, asynchronously and independently of other subjects. We use Qualtrics. Randomization by the software is independent across subjects.

Before the first round starts, participants are asked to estimate the probability that they are facing the high payoff stock. The correct answer is 50% as participants were told that the computer randomly selects the stock type with equal probability. Participants next allocate an endowment of 100 experimental currency units (ECU) between the stock and cash.[10] A snapshot of the initial round allocation screen is displayed in Panel I of Appendix A. The computer then randomly generates an outcome consistent with the distribution of the stock given its true type, and participants receive a report of the results of their investment round, i.e., whether the stock doubled or halved, as well as the value of their winnings. Having observed whether the stock doubled or halved in the first round, participants are asked to again estimate the probability that the stock is of the high payoff type and to allocate 100 ECU between the stock and cash. This process repeats until six investment rounds are completed. At the end of each round, participants are shown a history of their probability estimates, their stock allocations, the stock outcomes (whether it doubled or halved) and their winnings from each prior round. Panel II of Appendix A depicts a screen capture of what a subject might see after round three of a trial.

To encourage attentiveness, the experiment incorporates prompts asking participants if they are sure of their decisions whenever they appear to violate a monotonicity condition. For example, a prompt appears any time a stock outcome "halves" but a participant *increases* either the allocation to the stock or the estimated probability that it is a high payoff stock.

We now describe the full sequence of events (and trials). At the start of the experiment, participants are told that they will be taking part in an experiment in decision-making, and their main task will be to choose how to allocate an investment of 100 ECU between a risky stock and cash in each of a set of rounds. Participants are told that the total payout they can expect from the experiment consists of a US$7 participation fee, plus the total stock and cash payoff

[10] 100 ECU is equivalent to US$10.

from one randomly selected non-practice round, plus US$1 if the stock-type probability estimate made by the participant in the randomly selected round is within 5% of the true probability. Participants are also told that, given the stock's history, there is an objectively true probability that the stock is the high payoff type.

Each subject participates in seven trials divided into four blocks. The first block consists of a single practice trial and serves to familiarize participants with the process. The second block consists of two trials that set a baseline we term the Neutral treatment. A subject then participates in two additional treated blocks, each consisting of two trials: a Positive treatment block and a Negative treatment block (in randomly determined order). Because each trial consists of six investment decisions and probability elicitations, excluding the practice trial, we collect 36 observations of allocation decisions and 36 observations of likelihood perceptions, per subject.

Treatment proceeds as follows. Once participants complete the Neutral block, we elicit their social preferences by asking them to rank six social issues in order of importance. The six issues are: animal welfare, environment, refugees, poverty, human trafficking and gender discrimination. Participants then view a screen that describes two nonprofit organizations related to their top ranked social issue. They are asked to select one of the two nonprofit organizations to link to their trading outcomes. This process is repeated for the second-ranked social issue. Panel III of Appendix A depicts a snapshot of the social issue decision screens.

Participants subsequently proceed to either the Positive or Negative treatment block (the assignment is random). In the Positive block, participants are told that an amount of money equal to their stock payoff *would be donated* to a selected nonprofit by the experimenters. In the Negative block, they are told that an amount of money equal to their stock payoff *would be* *deducted* from the selected nonprofit's donation account.[11,12] It is important to note, as emphasized to the participants, that the amounts donated to (or deducted from) a selected nonprofit would not affect the participant's own gains during the experiment. The nonprofit remains fixed for both trials of a given block, but changes across treatment blocks. Panels IV and V of Appendix A show the instructions for the Positive and Negative blocks, respectively. Also shown in Panels IV and V of Appendix A is that during each trial, participants receive a report of the amounts to be potentially gained by, or deducted from, the nonprofit in past investment rounds (this can be compared with the feedback provided in the Neutral block – see Panel B).

The six non-practice trials in the experiment are payoff-equivalent for the subjects regardless of treatment. The only difference across blocks is that investment decisions may determine whether a nonprofit to which the subject exhibits some affinity gains (losses) money in the Positive (Negative) treatments. This allows us to examine the causal impact of treatment on likelihood perceptions and willingness to invest (i.e., preferences) given a likelihood perception.

2.2 Description of subjects

We recruited 160 participants from the University of Texas at Austin (62 identified as male, 97 as female, one did not identify themselves), receiving an average of $18.3 for their participation and choices (range of $12-$28, including a show-up fee of $7). The age of the participants

[11] We randomize whether the first or second ranked social issue's non-profit is assigned to the Positive or Negative block.

[12] For each subject, the randomly drawn payment round may be from the positive or negative condition, and associated with one of the subjects' two chosen nonprofits. We therefore aggregate the specified donations from each of the participants' outcomes across the non-profits, netting the positive and negative payments. For example, Rainforest Alliance was the non-profit associated with the payment round for 12 participants. The payment amounts were -$3.90 and $49.00 from the negative and positive blocks, respectively. Therefore, the final payment made to the Rainforest Alliance by the experimenters was $45.10. The net payment was negative for four of the twelve non-profits, and these payments were taken to be zero. Receipts for all payments were uploaded to a Dropbox folder and subjects were subsequently emailed a link to this folder.

ranged from 18 to 34 with a median of 20. This age group merits particular attention because of prevailing interest in the potential effects of millennials on asset markets.[13] During our study period millennials ranged in age from 23 to mid-30s. Most participants were students at the school, with 50 enrolled in business-related degrees, 39 in natural sciences, 19 in medicine, 16 in engineering, 10 in social sciences, and the remainder in arts/humanities, law, nursing, mathematics and communication. Somewhat surprisingly, demographic variables were not associated with the main effects we identify in the following section, and we omit them from our reported analyses.

One way of gauging subjects' comprehension and cognitive abilities around probability estimates is to ask whether their reported probability is 50% before observing any signals. Indeed, in 88% of trials subjects report the correct first-round probability, with this fraction being stable across treatments. Alternatively, 86% of subjects responded correctly on the first round always or almost always.

3. Empirical results

3.1. Overview

We focus on two variables, the participants' preferences for investment allocations and their beliefs, that is, the participants' allocations (in ECU) to the stock investment and their probability estimates of the stock being of the high payoff type. To analyze the treatment effect on each of these variables we begin with a simple comparison of their average levels across the three treatments (Negative, Neutral, and Positive). This is a very conservative use of the data as it produces a single observation for each subject-treatment, thus treating responses within it

[13] In particular, the increasing availability of RI products are often viewed as a response to anticipated demand from millennials, who are expected to receive large transfers of wealth from baby boomers in coming years.See, for example, https://pewrsr.ch/2Op4i3b; https://go.ey.com/2XvjCiP; and https://bit.ly/2O1r5mS.

as being perfectly correlated, and it lacks any of the controls that will be useful in isolating the treatment effects.

Table I: Stock Investment Allocations and Probability Estimates
This table reports the means, standard deviations, and ranges of the participants' choices regarding the stock investment allocations in each of the conditions and their estimates of the probability that they were facing a high payoff stock. The table also includes a test for whether the allocations and estimates are different across treatments, and a test for the effect asymmetry by calculating the ratio of allocation between the neutral minus negative treatments and positive minus neutral treatments ("asymmetry multiplier"). All *t*-test are based on matched sample at the subject-treatment level; *p*-values are reported.

Panel A: Stock allocation in ECU

	Negative (N=160)	Neutral (N=160)	Positive (N=160)	Total (N=480)
Mean (SD)	28.104 (17.817)	36.651 (20.557)	39.044 (21.739)	34.600 (20.606)
Range	0.000 - 100.000	4.167 - 95.833	2.833 - 100.000	0.000 - 100.000

Panel B: *p*-values from matched sample *t*-tests on stock allocation

	Negative=Neutral	Negative =Positive	Positive = Neutral
Probability	0.0%	0.0%	7.1%

Panel C: Asymmetry in Stock allocation ECU

	Neutral - Negative (N=160)	Positive - Neutral (N=160)	Asymmetry Multiplier	Test: Multiplier = 1 (prob)
Mean (SD)	8.547 (18.570)	2.392 (16.630)	3.5727	0.0081

Panel D: Probability estimates

	Negative (N=160)	Neutral (N=160)	Positive (N=160)	Total (N=480)
Mean (SD)	46.713 (13.039)	48.976 (11.492)	48.618 (12.327)	48.102 (12.316)
Range	9.667 - 75.917	5.833 - 77.417	8.833 - 81.667	5.833 - 81.667

Panel E: *p*-values from matched sample *t*-tests on probability estimates

	Negative=Neutral	Negative =Positive	Positive = Neutral
Probability	8.8%	12.3%	76.4%

In Table I we report the participants' allocations in Panel A, showing the differences across treatment conditions in the allocations to the risky stock. On average, subjects allocated 28.1, 36.7, and 39.0 (all out of 100 ECU) in the Negative, Neutral, and Positive treatments, respectively. Thus, relative to the Neutral treatment, the average allocation to the stock is 23% lower in the Negative treatment, but only 6% higher in the Positive treatment. Moreover, the difference between the Neutral and Negative condition is statistically significant at the 1% level, while the difference between the Neutral and Positive condition is only marginally significant at the 10% level (Panel B).

The magnitude of the deviation from the average Neutral treatment allocation is more than three and half times larger for "Neutral minus Negative" than for "Positive minus Neutral" and this "asymmetry multiplier" is significantly different from one at the 1% level (see Panel C). The study is designed so that the magnitude of the positive stimulus in the Positive treatment is, on average, the same as the magnitude of the negative stimulus in the Negative treatment. The profound difference in subjects' reactions potentially reflects an asymmetry in how the valence of these stimuli are perceived. For the sake of comparison, it is worth noting that experimental studies of loss aversion document an asymmetry multiplier that varies between 1.43 and 4.8 (Abdellaoui et al., 2007). Consistent with our effect being causally driven by considerations of externalities, we find that allocations are more affected by the treatment when it is randomly linked with a cause that the subject ranked first, compared to when it ranked second (27.7 vs. 30.5 ECU in the Negative treatment, 41.0 vs. 37.9 in the Positive treatment).

We also report the average level of the participants' estimates of the probability that the stock is the high payoff type. In comparing Panels D and A it is evident that the treatment effects are not as strong for probability estimates as for allocations to the risky stock. Probability estimates are similar at 46.7%, 49.0%, and 48.6% for the Negative, Neutral, and Positive treatments, respectively. Testing for the differences in probability estimates across the treatments suggests that only the Negative and the Neutral treatments are (marginally) significantly different (Panel E). There appears to be no significant difference between the probability estimates for the Positive and the Neutral treatments. The magnitude of probability differences between Negative and Positive (or Neutral) treatments are roughly consistent with the differences found in Kuhnen (2015) across probability estimates elicited in treatments of loss versus gain frames.

Table I provides a first glimpse of results that turn out to be robust in the experimental data: Subjects' allocation choices and probability estimates are far more sensitive to the Negative than the Positive treatment. That is, subjects in the Negative treatment sacrifice their own financial gain in the experiment to keep from harming charities that they selected. Moreover, the sensitivity to the Negative treatment is more pronounced for allocation choices than for probability estimates. It seems quite surprising that the response to Positive externalities is relatively muted. Put differently, subjects in the Positive treatment do not appear especially eager to take on greater risk, relative to the Neutral treatment, in order to raise the payoff of their chosen nonprofit. This is puzzling because charitable giving is both prevalent in practice and in experimental economics (e.g., see the meta-analysis of the Dictator Game in Engel, 2011). Consider that, in our design, *any* increase in allocation to the risky asset in the Positive treatment makes the associated nonprofit strictly better off.[14] Conventional wisdom, and intuition, might suggest that a significant number of subjects would therefore allocate more to the stock in the Positive treatment than they might otherwise. The fact that they do not signals that subjects respond to the investment context in a manner that is different from conventional charitable giving.[15] We interpret this as evidence that the presence of risk masks the charity giving context and creates a different frame in the minds of subjects. The finding that subjects are sensitive to the imposition of negative externalities is remarkable precisely because, in the same context, they are nearly indifferent to charitable giving.

[14] Even if subjects may not recognize this fact in the abstract, they can perceive it in the report we provide (e.g., Panel D, Appendix I) on how their allocations in the Positive treatment translate into outcomes for the nonprofit.
[15] List (2007), for instance, provides examples of contexts that can greatly influence pure charitable-giving motives via the Dictator Game.

3.2 Allocations

Although the allocations are affected by subjective probabilities, the summary statistics provide evidence that RI considerations are affecting allocations to the stock above and beyond what can be easily explained by shifts in beliefs alone. To help disentangle the effects, we focus on the allocation decision given subjects' likelihood perceptions (i.e., preferences), and later separately examine the effects of RI on probability estimates (i.e., beliefs).

Before applying a more structured approach to the data, we filter for weakly rational behavior on the part of the participants. By this we mean that a subject in the Neutral treatment should, *on average*, exhibit a weakly positive relationship between actual and subjective (i.e., estimated) probabilities that the stock is high payoff, and a weakly positive relationship between subjective probabilities and stock allocations. We interpret violations of these conditions in the Neutral treatment to signify lack of engagement or confusion about the basic experimental tasks. We test for the conditions, at the subject level, using linear regression. Of the 160 subjects, 35 were dropped because they violated one or both of these requirements. It is important to note that this filter is only applied through the Neutral treatment and thus does not implicitly condition on behavior in the main treatment cells of interest – the Positive and Negative conditions. We proceed by analyzing the data from the remaining 125 subjects though it bears emphasizing that we obtain qualitatively similar results without this filter.[16]

Table I, Panel D, indicates that subjective probabilities in the Negative treatment are below those in the Neutral or Positive treatments. That would be sufficient to predict lower allocations in the Negative treatment. To control for the impact on probability assessment and examine the treatment effects on allocation separately from their effects on beliefs, we pool

[16] Together with the subject prompts (in the event of monotonicity violations), the filter effectively acts as a control for comprehension. The fact that the results are largely robust to explicitly controlling for these effects is reassuring.

allocation observations across all rounds based on subjects' *reported* (i.e., subjective) probability bins as depicted in Figure 1.

A number of suggestive patterns emerge from the plot. First, allocations in the Negative treatment are lower across almost the entire range of subjective probabilities compared with the Neutral or Positive treatments. Second, the effect does not appear to be uniform – instead, the difference between the allocations in the Negative treatment and the other treatments appears to increase as subjective probabilities increase. This is especially striking when plotting differences from the Neutral treatment (Panel B) where one sees both the asymmetry in response to the treatments as well as the subjective probability dependence. Finally, there appears to be only a marginal treatment effect on allocations when comparing the Positive condition to the Neutral one.[17] To further quantify these observations, we regress allocations in each round on treatment dummies, the reported subjective probabilities, and interactions between treatment dummies and the reported probabilities. All regressions include subject fixed effects to control for heterogeneity in average allocations across subjects.

[17] The average Neutral treatment allocation per bin is greater than that of the Positive treatment in only two of eleven plotted bins. Under a 50-50 null (consistent with indifference) between the bin statistics, this is associated with a two-sided p-value of 6.6%.

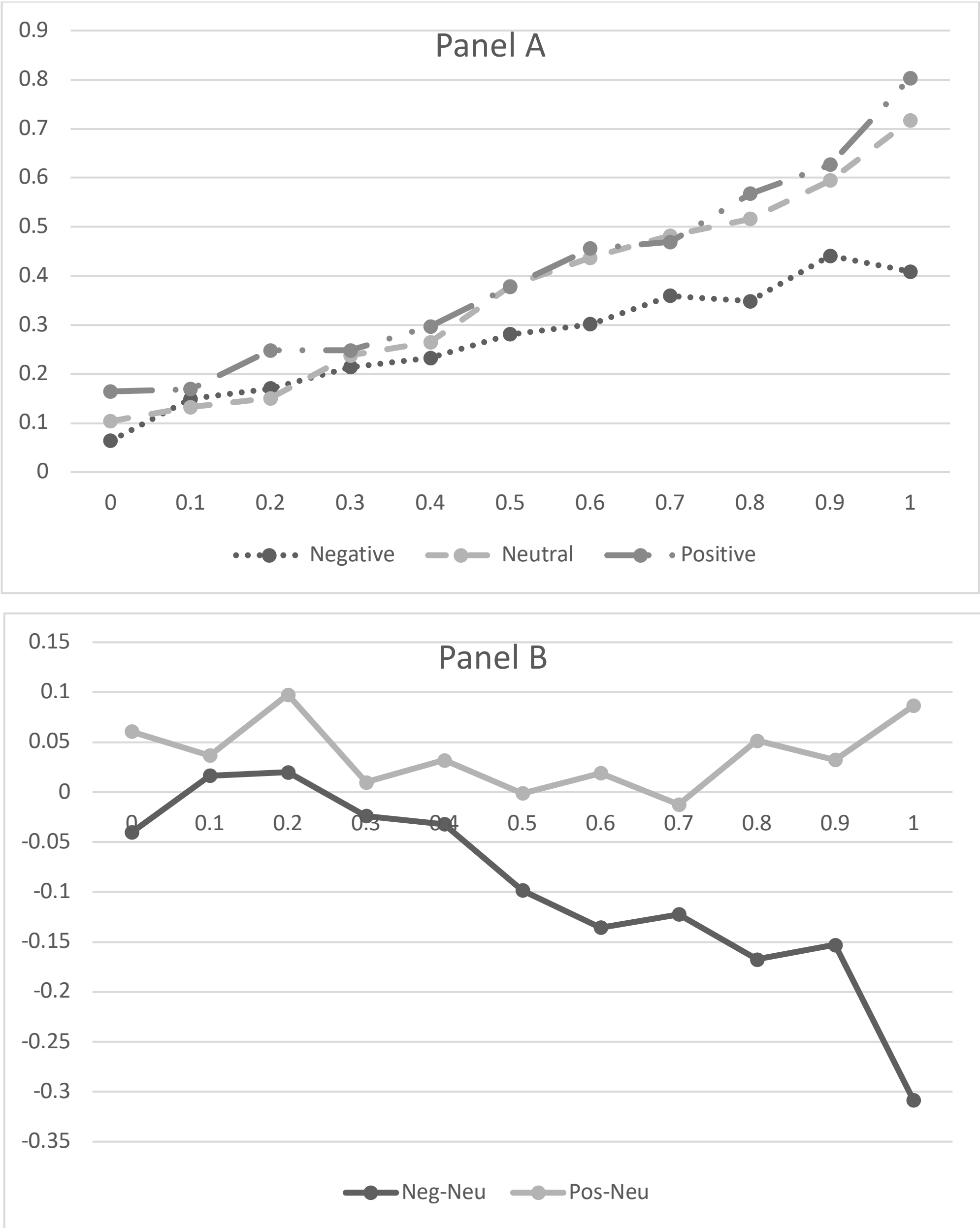


Figure 1: Treatment effects on allocations

This figure shows the treatment effects on allocation (y-axis), scaled to 0.0-1.0, against subjective probability estimates that the stock is of high type (x-axis). We group observations based on the subjects' reported rather than objective probabilities to control for allocation differences that may arise from different assessments of probabilities. Panel A plots the average allocation across the three treatments while Panel B plots the average difference in allocation in the positive and negative treatments, both relative to the neutral treatment.

Table II confirms the patterns reported in Table I and Figure 1. In Column (1), allocations in the Negative treatment, which are the baseline, are on average a highly significant 9.6 ECUs lower than in the Neutral treatment. The increase in allocation observed in the Positive treatment is 1.6 (=11.2 - 9.6) ECUs higher than in the Neutral treatment, which is barely significant at the 5% level. The results reported in Column (2) demonstrate that, while there is a strong treatment effect, it does not appear to have a significant constant component (the un-interacted treatment dummies). On the other hand, we find that the response of allocations to probabilities is much lower in the Negative treatment relative to the Neutral treatment, 0.381 vs. 0.589, but there is no significant difference between the Neutral and Positive treatments. This is consistent with a more pronounced allocation reduction in the Negative treatment, relative to the others, as subjective probabilities increase.

Table II: Stock allocations and probabilities

The table reports regressions of round-by-round percentage allocations to the stock on reported subjective probabilities ("Prob"), treatment dummies, and interactions between reported probabilities and the treatment dummies. The Negative treatment is the baseline. Standard errors are in parentheses. All regressions include subject fixed effects. The analysis is limited to subjects exhibiting weak rationality in the Neutral treatment (see text). The asymmetry multiplier is calculated as the ratio between the allocation in the neutral-negative treatment and the positive-neutral treatment (in column 2, the ratio is evaluated at the median probability of 50%).

	(1)	(2)
Prob		0.381***
		(0.023)
Neutral Block Dummy	0.096***	-0.012
	(0.008)	(0.017)
Positive Block Dummy	0.112***	0.018
	(0.008)	(0.017)
Prob x Neutral Block Dummy		0.208***
		(0.032)
Prob x Positive Block Dummy		0.182***
		(0.033)
Observations	4,500	4,500
R^2	0.369	0.527
Asymmetry Multiplier	6.00	6.27
Test of asymmetry multiplier	0.000	0.000

The greater number of observations in the analysis of Table II permits a more precise estimate of the asymmetry multiplier (the ratio of the magnitudes of the deviation from the

Neutral treatment). Using estimates from the regression parameters, and evaluating it at the (roughly) median subjective probability of 50%, the multiplier is close to six – corresponding to substantial statistical and economic significance.

The results from Column (2) in Table II may be interpreted as evidence that the channel through which the negative treatment primarily impacts subjects' preferences is their sensitivity of allocation to subjective probability: The greater the probability of a high stock payoff, the greater the difference between the allocation between the Negative and the other treatments. Note that a significant "charitable-giving" motive implies a significant Positive block dummy coefficient. This is because any additional amount allocated to the stock in the Positive treatment results in some distribution to the nonprofit (even when the subjective probability is zero that the stock is high-type). That said, consistent with the absence of a charitable-giving motive, allocations in the Positive and Neutral treatments are also similar in their sensitivity to the subjective probability of a high payoff.

3.3 *External validity and survey evidence*

The positive treatment in our experiment mimics certain RI objectives in that it creates "good" externalities. Although the negative treatment does not directly create harm, it can be viewed as creating less good. However, this may raise questions as to whether our experiment measures an asymmetry that is meaningful in practice. Given the institutional restrictions that rule out a design that could be plausibly linked to actual harm creation, we conduct a separate survey to test whether our results *understate* the strength of asymmetries. That is, we hypothesize that subjects' negative reactions to funding a group that advocates against environmental or social causes would be stronger than their negative reactions to taking funds away from a group that advocates for these causes. In this subsection, we report on the results of a separate survey that strongly support this hypothesis.

The survey was run on Prolific, pooling from the general US adult population. To match subjects' age and gender demographics in the main experiment, we specified that participants be between the ages of 18-25 and requested an equal proportion of males and females. In all, 108 participants responded to the survey.

To mirror the design of the main experiment, at the start of the survey, participants are asked to rank the six available causes, which were identical to the ones used in the main experiment. For their highest ranked cause, each subject is asked to make a hypothetical choice between two negative effects: (i) providing funds to a firm that lobbies against their chosen cause (i.e., harm creation) and (ii) cutting funds from a charity that lobbies for their chosen cause. Subjects are also given the option to indicate that both choices are equally bad. An example of this choice for a social cause concerned with poverty is provided in Appendix B.

Table III summarizes the main survey results. Among the 108 responders, 36% preferred the choice of cutting funding to the initiative that would promote their cause and only 11% preferred the choice of contributing to the initiative that would harm their cause. The remaining 53% were indifferent, that is, they indicated that they thought both choices were equally bad. Using a simple Chi-squared test, at a 1% level we can reject the null that choices are uniformly distributed across the three options provided. This allows us to reject the null hypothesis that most respondents randomized their choices. Likewise, we can reject the null that respondents were equally disposed towards reducing good versus creating harm: Three times more subjects preferred to reduce good than create harm. Viewed another way, out of the respondents who indicated a preference, 76% selected the reducing good choice over the creating harm choice.

At the end of the survey, participants were presented with a few exit questions, including a self-evaluation of their level of engagement (on a scale of 1-10). This provides an indirect proxy for their level of interest in these RI issues overall. We divide the subject pool into those that assessed their engagement as “10” (61%) and the remainder (39%). While the maximally engaged group exhibits a far greater propensity to reduce good than create harm, remaining subjects still exhibited a significant preference in the same direction.[18]

Table III: External validity survey

The table reports survey results testing whether subjects would choose to “reduce good” by cutting funding from a charity lobbying to ameliorate a social issue they rank as important, “create harm” by donating to a firm lobbying to exacerbate the social issue, or are indifferent between cutting and donating. The first Chi2 column reports the *p*-value for a Chi-squared test for uniformity of responses (i.e., randomization). The second Chi2 column reports the *p*-value for a test of the null that the propensity to cut is the same as to donate in the population. The subsample of subjects who are “Maximally engaged” assessed their engagement in the survey as “10 out of 10”.

	Cut	Donate	Indifferent	Total	Chi2(Uniform)	Chi2(Cut=Donate)
Total	39	12	57	108	0	0
Maximally engaged	24	5	37	66	0	0
Others	15	7	20	42	0.0464	0.088

The results of the survey strongly support the thesis that the subjects in the main experiment would have exhibited even greater aversion to the negative treatment if it had incorporated harm creation in the design. In other words, our estimates from the main experiment likely *understate* the true asymmetry in RI preferences.

3.4 Heterogeneity of treatment effects

The results so far focus on averaging treatment effects across subjects. An interpretation of the results in Table II in terms of an impact on individual preferences is, at best, indirect because the analysis relies on the average treatment effect across all subjects. In other words, Table II reports aggregated choices but is not directly informative about the behavior of the average individual. The primary focus of this subsection is to quantify the heterogeneity of RI preferences in the subject pool. We investigate whether the average treatment effect is broad-

[18] We also included a task comprehension question at the end of the survey. Of the 108 participants, all but 2 answered the comprehension question correctly. Moreover, 87%/92% of subjects rated the survey as being very engaging/clear (giving it a score of 8 or more out of 10).

based or attributable to a few subjects with very strong social preferences. Arguably, a prevalent effect can be more compellingly extrapolated to the general population. In addition, the prevalence and strength of the effects can be potentially useful in calibrating models that incorporate social preferences and which often include more than one investor type.

We tackle this question using both a semi-parametric (main analysis) and fully parametric (appendix) approach. The advantage of the first approach is its simplicity and the weak assumptions required. The second approach is more involved but, arguably, makes better use of the data and better addresses potential bias arising from the combination of experimental noise and the constraints on stock allocation.[19] The broad conclusion from both approaches is consistent in that we find no evidence of a cohort that dramatically shifts their allocation in the Positive treatment (relative to the Neutral treatment), while finding strong evidence for a substantial cohort that sharply cuts their stock allocations in the Negative treatment.

Because the focus of this subsection is on individual-level preferences, we create a simple measure of "excess" and "deficit" allocations in the Positive and Negative treatments, respectively, that is conditional on probability. To start, we average each subject's excess stock allocations in the Positive treatment relative to the Neutral treatment for the three separate states in which the objective probability that the stock is "high payoff" is 1/3, 1/2, or 2/3. The averaging is done to reduce within-subject noise. For the Negative treatment, for each subject, we instead calculate the average *deficit* allocation in the Negative treatment relative to the average Neutral treatment allocation (again, separately for each of the objective high payoff probability assessments, 1/3, 1/2, and 2/3). That is, in the Positive (Negative) treatment, for each subject and the three probability states, we calculate the average (negative) deviation from the Neutral treatment.

[19] Results of the more elaborate estimation procedure outlined in Appendix C can also be used to make more detailed and direct comparisons with predictions of theoretical models of individual RI behavior.

Next, for each of the six treatment-probability combinations, we fit the deviation data to a mixture of two normal distributions. The model assumes that subjects come from one of these two distributions, and thus estimates the means and standard deviations of the two distributions, as well as the fraction of subjects that are associated with each. We also do this for excess/deficit allocations after averaging deviations across all three probability scenarios. Intuitively, this procedure allows us to quantify, statistically, heterogeneity across subjects while reducing within-subject noise.

Table IV reports the means for the two normal distribution components and the share of subjects estimated to be associated with the first (lower variance) component. In each instance of model estimation the two variance components are significantly different, suggesting that the subject population consists of at least two behavioral types.[20] The component means in Table IV are reported as a percentage of the average Neutral treatment allocation for the corresponding probability assessment. Focusing first on the Positive treatment, we see that the two mean components cannot be statistically distinguished from each other or from zero, at the 5%-level, in any of the Table IV estimates. In other words, while one can statistically sort subjects into "high" and "low" variance types in their deviation from the Neutral treatment, the average deviation is statistically close to zero for both types.

The results are starkly different for the deficit allocation in the Negative treatment. Subjects are drawn either from a low-variance population that reduce their allocation by a modest average of roughly 10% relative to the Neutral treatment, or from a high-variance distribution that reduce their allocation by a whopping average of 60% across the objective probability scenarios. The share of subjects who react more strongly to the Negative treatment increases with the probability that the stock type is high, and consists of roughly half of all

[20] We do not report the respective variance components estimates in Table IV as they are not germane to the question we explore. The first component normal distribution standard deviation estimates range from 4% to 12% while those of the second range from 22% to 43%.

subjects when this probability is 2/3. In the latter case, both the economic and statistical significance is very high.

Table IV: Subject Heterogeneity

The table reports estimates of a mixed-model to subjects' allocation deviation from their Neutral treatment response in three separate objective probability scenarios (as well as the scenarios' average). In each of the objective probability scenarios, we calculate for each subject the average difference between their stock allocation in the Positive and Neutral treatments. The resulting subject-level distribution is then fit to a mixture of two normal distributions. The estimated parameters in the estimate are the two variances, two means, and the mixture parameter (the proportion of subjects estimated to belong to the distribution with the lower estimated variance). For the Negative treatment, we instead calculate, for each subject, the average difference between the Neutral and Negative treatment allocation. In the table, the 1st (2nd) component refers to the estimated normal distribution with the lower (higher) variance. We only report the means of the two components and the mixture parameter (1st component share) and they are reported as a percentage of the average Neutral treatment allocation across all subjects. The bottom panel reports the mixture model fit when averaging over allocation difference in all three probability scenarios.

	Positive treatment		**Negative treatment**	
Variable	Estimate	s.e.	Estimate	s.e.
Objective Probability = 1/3				
1st component mean	1.3%	2.8%	14.3%	6.0%
2nd component mean	26.0%	15.5%	56.6%	42.0%
1st component share	48.40%	7.88%	81.90%	8.45%
Objective Probability = 1/2				
1st component mean	0.5%	2.8%	8.9%	4.3%
2nd component mean	13.5%	20.7%	62.4%	21.4%
1st component share	81.90%	10.16%	78.10%	11.51%
Objective Probability = 2/3				
1st component mean	4.5%	3.1%	4.7%	2.6%
2nd component mean	0.6%	7.0%	57.9%	12.5%
1st component share	42.70%	11.27%	51.00%	7.52%
Objective Probability = 1/3, 1/2, or 2/3				
1st component mean	1.73%	2.51%	10.71%	3.85%
2nd component mean	31.63%	25.72%	65.91%	27.25%
1st component share	88.70%	6.33%	75.00%	14.92%

The fit to the mixed model also allows us to estimate, from Bayes' Law, the probability that a specific subject's reaction is drawn from the first (low variance) component. Interestingly, the correlation of this probability across treatments is modest at about 33% when averaging over all three probability scenarios. By contrast, the *within treatment* correlation of this probability between the 1/3, 1/2, and 2/3 scenarios is high at about 80% for each of the Positive and Negative treatments. In other words, there is substantial subject consistency in behavior *within* treatments but not as much *across* treatments. This supports the notion that attitudes towards the Positive and Negative frames are distinct even within person.

Finally, we consider the asymmetry of the treatment effect within subjects. For each subject, we subtract the Positive treatment effect, averaged across the three objective probability scenarios, from the corresponding averaged Negative treatment effect. When this difference is positive, the subject exhibits a stronger Negative than Positive treatment effect. Figure 2 plots the density of this asymmetry measure across the subjects. Roughly 2/3 of the subjects are characterized by more pronounced Negative than Positive treatment effects. The average and median asymmetry is 6.9 and 5.4 ECUs, respectively, which is roughly 20% of the corresponding Neutral treatment stock allocation (averaged across the 1/3, 1/2, and 2/3 scenarios). Within subject asymmetry of the treatment effects can be seen to be prevalent and not driven by a few outlier subjects.

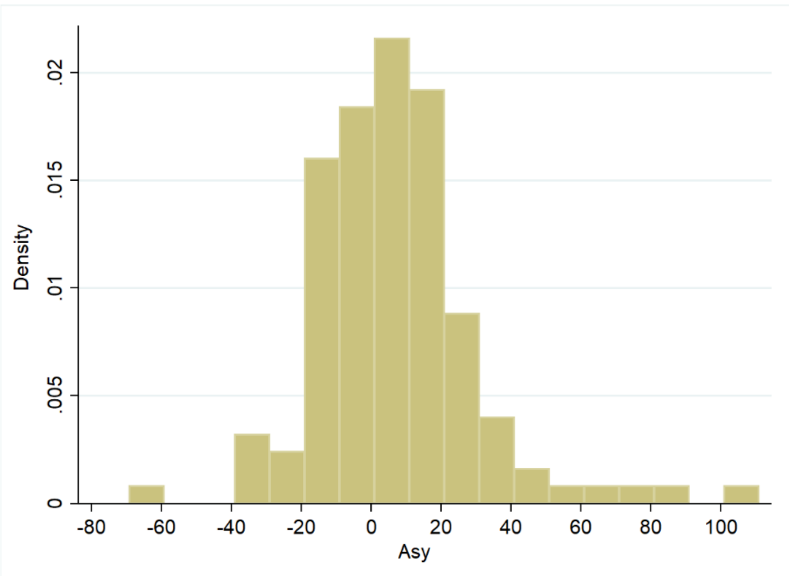


Figure 2: Asymmetry in subject reaction to the Negative and Positive Treatments
A positive number corresponds to a larger stock allocation deficit in the Negative treatment, relative to the Neutral treatment, than excess stock allocation in the Positive treatment.

Theory and intuition suggest that the magnitude and prevalence of the effects we find across subjects, if representative of the population, will impact asset prices (e.g., Heinkel, Kraus and Zechner, 2001; Zerbib, 2022). The relatively weak Positive treatment results and strong Negative treatment results (i.e., the asymmetry) appear to be reflected in the empirical literature reviewed in Section 4.2, where the evidence for an angel stock valuation premium is mixed while the evidence for a sin stock discounts appears consistent.

While the analysis of heterogeneity presented above has the virtue of being relatively simple, it is subject to several potential shortcomings. First, in order to achieve enough statistical power at each probability level of the stock being "high payoff", we resort to using the objective rather than the reported probabilities. This also forces us to rely on only three objective probability level: 1/3, 1/2, and 2/3. Moreover, we effectively assume that the three probability levels are independent within subject. Second, within subject experimental noise may bias allocations because the possible allocations are bounded. In Appendix C we address these by estimating a subject-level stimulus-response model that incorporates experimental noise and all *reported* probability observations. This more elaborate approach confirms the findings described above: A substantial proportion of subjects exhibit asymmetric treatment effects, with a large negative treatment effect that increases with the probability that the stock is high type (see Figure C-1).

3.5 Subjective Probabilities

3.5.1 Probability estimates

We now turn to assessing the effect negative and positive externalities have on subjective probability estimates. The histogram in Figure 3 compares true (objective) versus subjective (reported) probabilities for various true probability bins. The Bayesian objective probability of the stock being of the high return type is calculated as follows: given a history of *n* doubling and *m* halving outcomes, it is $2^{n-m}/(2^{n-m}+1)$. Given the discrete nature of signals and the asymmetry of updating, in most rounds, the true probability will be one of 1/3, 1/2, and 2/3 (corresponding to $|n - m| \leq 1$). We therefore use five true probability bins: below 1/3, exactly 1/3, exactly 1/2, exactly 2/3, and above 2/3. The thick black bars indicate the expected true probability for the corresponding bin (exactly 1/3, 1/2, or 2/3 for the inner bins, and the expected value of the true probabilities in the outer bins). The bars depict the subjective probability for the different treatments.

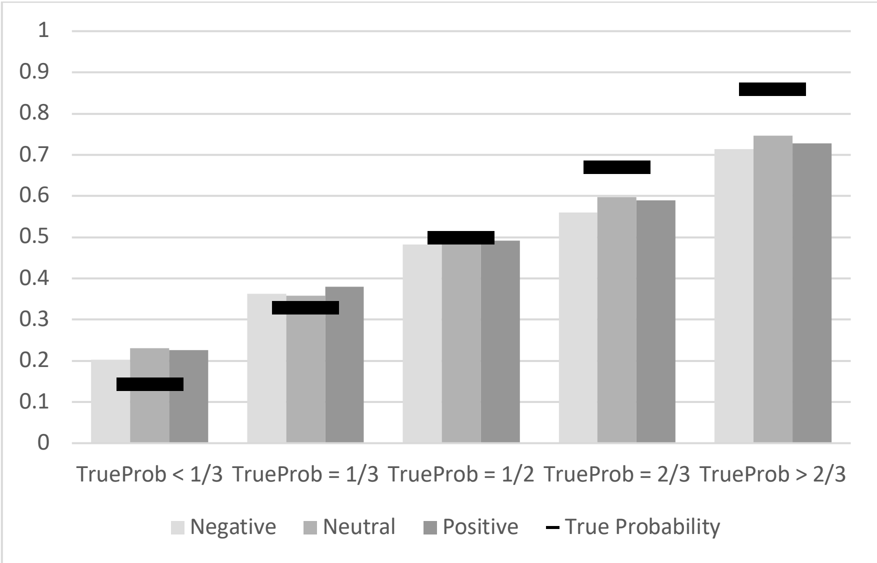


Figure 3: Objective and subjective probabilities
The figure depicts objective probabilities of the stock's type to subjective (reported) probabilities. The thick black bars indicate the expected true probability for the corresponding bin (exactly 1/3, 1/2, or 2/3 for the inner bins, and the expected value of the true probabilities in the outer bins). The bars depict the average subjective (reported) probability for the different treatments.

A couple of patterns emerge in Figure 3. First, consistent with a large prior literature (e.g., Tversky and Kahneman, 1992, Abdellaoui, et. al., 2011, and Kuhnen, 2015), we find that subjective probabilities are "shrunk" toward the unconditional prior of 1/2. That is, when objective probabilities are less (more) than 1/2, subjects' perception of probabilities are too high (low). Second, we find that subjective probabilities in the Negative condition tend to be lower than the probabilities in the Positive condition. This difference is around 2 percentage points across objective probability bins.

The first pattern noted above resembles the behavior of "subjective weights" in non-expected utility models, such as Cumulative Prospect Theory. In particular, Prelec (1998) suggests the following formulation, adapted to our multi-treatment setting:

$$Subjective\ weight = \exp\left(-\left(\delta + \delta_{Pos} + \delta_{Neg}\right) * \left(-\log(True\ Prob)\right)^{\gamma+\gamma_{Pos}+\gamma_{Neg}}\right). \quad (1)$$

In the specification above, the Neutral treatment is the baseline. One can roughly think of the δ's as level parameters, shifting subjective probabilities up or down relative to objective probabilities, and of the γ's as curvature parameters. The Bayesian null is $\delta = \gamma = 1$ and $\delta_{Pos} = \delta_{Neg} = \gamma_{Pos} = \gamma_{Neg} = 0$. We adopt this formulation and fit it to our stated beliefs, across all rounds.

We report in Figure 4 the estimated subjective probability function parameters across the three treatments, and plot the fitted treatment-dependent subjective probability as a function of objective probabilities. The subjective probability plots in the Positive and Neutral treatments are very close, while those of the Negative treatment are consistently below the other two, across the entire range of objective probabilities. The shift down, in the Negative treatment, appears to be similar across the range, consistent with the treatment effect on curvature being insignificant.

Variable	Estimate	s.e.
δ	0.9282	0.0113
γ	0.6329	0.0170
δ_{Neg}	0.04618	0.0164
γ_{Neg}	-0.0286	0.0238
δ_{Pos}	-0.0022	0.0159
γ_{Pos}	-0.0267	0.0237

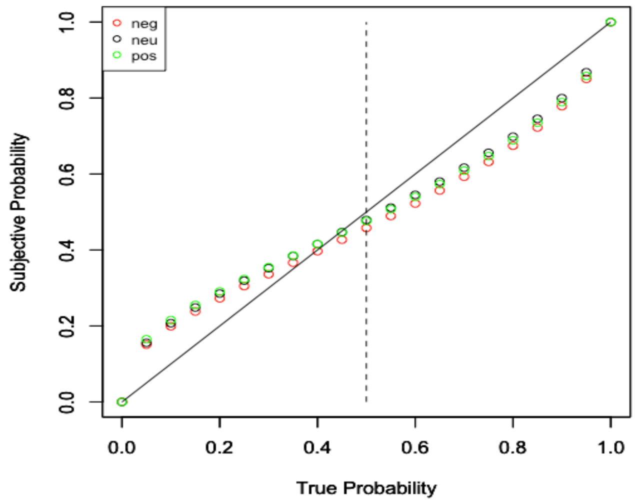


Figure 4: Cumulative probability function
This figure shows the maximum-likelihood estimation of equation (1) along with a plot that depicts the estimated parameters. Standard errors are in parentheses.

The asymmetry between the Negative and Positive treatments is borne out by the estimation results. There appears to be no statistically significant difference between the Positive and Neutral ones. However, both δ_{Neg} and γ_{Neg} are different from the corresponding Neutral and Positive treatment coefficients.

In Appendix D we employ the Mobius (2022) model to pin down the source of asymmetry between updating in the Negative versus the Positive/Neutral treatments. The analysis suggests that the probability assessment bias observed in the Negative treatment arises from lower sensitivity to "good news" about the stock's payoff distribution.

All in all, the belief results show modest, but surprising effects. Overall, we find no effect on beliefs or their updating in the Positive treatment. In contrast, we find muted response to positive information in the Negative treatment that is consistent with lower subjective probability estimates in this treatment over the high objective probability range.

Finding an effect on beliefs is surprising given that reported subjective probabilities, unlike stock allocations, have no externality on the linked cause; in fact, subjects are explicitly paid for accuracy. Second, the experiment is set up in a way that delinks the prospects of the stock from the cause and so beliefs should have remained constant even if the linked cause is one over which the subject has strong preferences. That said, our results on beliefs and updating echo the findings around at least two well-established behavioral biases – the halo effect and the confirmation bias. The halo effect suggests that positive or negative impressions carry over irrationally across domains in a way that is often unconscious and that can be affected by ESG policies (e.g., Nissbet and Wilson (1977), Hong et al (2019)). The confirmation bias suggests that individuals overweight evidence that is consistent with their priors (see Klayman (2008) for a survey). While, again, our findings are consistent with these biases, our contribution is to show that they are invoked only by the concerns for negative, but not for positive, externalities. The spillover effect of framing on beliefs is also consistent with

Kuhnen (2015) who finds that subjects' learning process – in this very task – differs based on whether outcomes are in the domain of gains or losses.

4. Further Discussion

Our main findings concern the asymmetric impact of negative RI externalities on investment preferences and belief formation, and on the distribution of sensitivities to RI externalities in our sample of subjects. These findings, in turn, separately contribute to different branches of the RI literature.

4.1. Pecuniary alignment

Because ESG rankings may correlate with perceived risks, such as future regulations (Stroebel and Wurgler, 2021; Krueger, Sautner and Starks, 2020), the question of whether RI represents an important departure from the traditional paradigm of balancing risk-return financial tradeoffs has been hotly debated. Investors who associate high ESG-rated stocks with higher expected returns or lower risk will allocate more capital to these investments, much as would investors who are motivated primarily by non-pecuniary motives.[21] While a number of studies attempt to establish non-pecuniary motives as an important aggregate investment channel, this is challenging to confirm in practice. Because our experimental design only varies non-pecuniary externalities across the three treatments, we are able to directly observe the presence of non-pecuniary preferences. The results show that non-pecuniary preferences are pervasive across subjects and economically large, which supports the broader claims of empirical studies pointing to the importance of such preferences and where controlling for financial outcomes is difficult or impossible (see references in Footnote 4).

[21] For discussions of the different types of RI investment motivations, see, for example, Edmans and Kacperczyk (2022) and Starks (2023).

4.2. Asymmetry in RI approaches and impact on stock returns

Our findings resonate with a number of empirical stylized facts, suggesting they may help provide the basis for microfounding a descriptive model of individual RI preferences that can also explain aggregate behavior. For example, our asymmetric results on preferences in the presence of negative versus positive externalities are consistent with the asymmetry found in studies of fund flows and stock market reactions which appear to be more strongly affected by negative RI events.[22] Similar to our results but under a different experimental setting, Chew and Li (2021) find strong evidence for asymmetry in sin stock aversion versus virtue stock affinity in their experimental study, which they model by appealing to the literature on source-dependent risk aversion.

Our results also conform to what we observe in practice in the RI market. Although much of RI has evolved to include positive ESG tilting, negative screening remains a pervasive strategy.[23] An often-followed model of this investment approach is the Norwegian sovereign wealth fund (the Government Pension Fund Global) which invests sustainably but also excludes companies that they believe do not meet their ethical norms.[24]

Our experimental design allows us to control for the magnitude of the externality while switching its sign. This is not generally possible in natural experiments, for example, where ordinal ESG rankings are used to compare RI alternatives. In addition to providing a "clean" causal channel for observed asymmetry, our experiment allows us to rule out the possibility

---

[22] See, for example, Krueger (2015); Bialkowski and Starks (2018); Hartzmark and Sussman (2019).
[23] For example, of the 86 U.S. investment managers who reported their screening technologies to the Principles of Responsible Investing, 91% use some type of negative screening strategy. In fact, 33% use only negative screening while 58% use negative screening combined with some form of positive screening. Similarly, a survey finds that 69% of managers use negative screening in their investment decisions (US SIF, 2020). We thank the PRI for providing us with 2018 data to calculate these statistics. The PRI has not reviewed the methodology applied to, use of, or conclusions drawn from this data.
[24] https://www.nbim.no/en/the-fund/responsible-investment/.

that the asymmetry exists because poorly ranked ESG alternatives are significantly worse from a measurable social perspective than highly ranked ESG alternatives.

Furthermore, RI models in which investors exhibit a pronounced preference for stocks that rate highly on an E, S or G dimension predict a valuation premium for such stocks (e.g., Pastor, Stambaugh, and Taylor, 2021). The empirical literature, however, lacks a consensus around this intuitive prediction.[25] Our findings of only a weak preference for investing in experimental stocks linked to positive externalities is consistent with the weaker evidence in the empirical literature for the performance of strategies that favor investment in ESG. Our stronger findings concerning negative externalities are consistent with the clearer empirical evidence of shunned stock undervaluation (e.g., Hong and Kacperczyk, 2009; Statman and Glushkov, 2009). Thus, our findings provide a way to interpret the array of evidence documented in a large and growing literature and should help in the development of new theories.

4.3. The role of distorted beliefs

Importantly, our study points to another potential (albeit weaker) channel impacting asset markets by RI investors. If the ability of individuals to infer the likelihood of outcomes is impacted by negative externalities as our results imply, asset prices will be affected beyond what is suggested in earlier theoretical work that focuses only on the impact of tastes on allocations. If a sufficiently large proportion of RI investors in the economy deviates from Bayesian updating because of RI sensitivities then additional distortions could arise from the effective presence of "pessimistic" investors. Because not all models incorporate non-Bayesian

[25] See, for example, Statman and Glushkov (2009), Edmans (2011), Humphrey, Lee, and Shen (2012), Bansal, Wu, and Yaron (2022), Hwang, Titman, and Wang (2022) and Liang, Sun, and Teo (2022)

beliefs, it seems useful to disentangle the impact of RI on preferences versus beliefs. Our results provide empirical guidance on how this can be done.

4.4. Distribution of RI preferences and theoretical RI models

Our results, particularly the results regarding the effects of negative RI information on investors' asset allocations and beliefs, have important implications for financial markets. If the percentage of RI investors grows in the economy as expected by many, asset prices are likely to be affected by their allocation choices, which has been shown theoretically and suggested empirically. For example, Heinkel, Kraus and Zechner (2001) and Luo and Balvers (2017) demonstrate how shunning sin stocks would have the effect of driving these stocks' prices lower, which has empirical support in Hong and Kacperczyk (2009), Statman and Glushkov (2009), as well as Chava (2014). Similarly, Fama and French (2007) provide a simple theoretical framework to demonstrate that investor tastes, such as tastes for responsible investing, can distort pricing in asset markets. They show that these distortions in prices could be large under certain circumstances: when investors with particular tastes represent a substantial fraction of invested wealth; when the investors have such tastes for a wide range of assets; when investors' positions vary quite a bit from the market portfolio; and when the returns on the investors' underweighted assets are not highly correlated with the returns on their over-weighted assets. In other words, it is plausible to expect an impact on asset prices when responsible investors represent a substantial percentage of investors in the market.

As many as half of our subject pool would substantially reduce their allocation to an asset with a negative RI association. The magnitude of such RI preferences, when extrapolated to the population, should be sufficient to significantly impact asset prices. Moreover, the asymmetry we find in RI preferences is consistent with only a subset of existing theoretical models (as we point out in Footnote 4).

4.5. Additional experimental studies

Three recent studies, all of which are based on Dutch investors, examine investors' choices of responsible investing strategies. Riedl and Smeets (2017) conclude that intrinsic social preferences and social signaling are the investors' primary motivations and that while financial motivations enter into the decision making, they play a relatively minor role. Brodback, Guenster and Mezger (2019) find a positive link between altruistic values and the relative importance of social responsibility to investors. They also find that the link strengthens under certain conditions: when individuals believe their investments can make a social or environmental impact or when they feel moral obligations regarding their investments. Lastly, Bauer, Rouf and Smeets (2021) conduct a field experiment in which Dutch pension participants are allowed to vote on whether the pension system should follow three or four of the United Nations Sustainable Development Goals. They conclude that the choice of 66% of the participants to follow more of the goals, i.e., engage in more responsible investing activities, is based on nonfinancial rather than financial considerations. Whereas these papers seek to answer the question of *why* investors select into being RI investors, our work investigates *how* the RI information is incorporated into investors' decisions.

Our analysis is also complementary to several contemporaneous experimental studies, but the focus and consequently, the experimental design, exhibit key differences. Bonnefon et al. (2022) examine the private valuation assigned by MTurk subjects to direct giving to (or taking from) charities, and find it is roughly linear in the small stakes considered. Moreover, private valuations do not significantly depend on whether a subject is pivotal to the giving (i.e., whether the amount the charity receives depends on actions taken by the subject). Our design is fundamentally different in that it incorporates and examines dimensions of quantity, uncertainty, and learning linked to the RI decision with a specific focus on asymmetry. On the other hand, we do not test for a difference between pivotal and non-pivotal treatments. Whereas both our paper and Bonnefon et al. provide strong evidence that tastes matter in evaluating RI, their paper focuses on the difference in pivotal vs. non-pivotal effects and appears to find no evidence for the strong asymmetric results discovered in our setting.[26] In their paper, there is no room for updating so they cannot separately assess the effect of the treatment on beliefs vs. allocations.

Brodback, Guenster and Pouget (2020) also conduct an experiment with charitable donations tied to a financial investment, in this case an initial public offering. They find that individuals have a price premium for social responsibility that also depends on the asset's financial performance. Their focus is on the pricing of the social benefit in combination with the financial performance while our focus is on how the externalities affect beliefs and allocations. Likewise, Heeb et. al. (2021) examine investors' willingness to pay for sustainable investment in an experimental setting. They focus on positive externalities and how willingness-to-pay changes with impact and the choice set.

Finally, we point out that a vast literature exists on "other-regarding behavior," mostly focused on strategic choice problems.[27] Although gain-loss asymmetry, introduced in Kahneman and Tversky (1979), is one of the most influential and persistent stylized facts in human decision making, the evidence we find for its social preference manifestation appears to be new.

5. Conclusions

In this paper we employ an experimental setting to study how social externalities influence individuals' investment decisions and belief formation. We find that linking individual

[26] The significantly smaller stakes employed in Bonnefon et al. (2022) for both subjects and charities might serve to mask a difference between the pivotal and non-pivotal treatments, or an asymmetry between the impact of negative versus positive payoffs to the charity.

[27] See Cooper and Kagel (2016) for a review.

investment outcomes to social externalities significantly affects individuals' allocations between a risky asset and cash as well as their subjective beliefs over the investment outcomes.

Beyond disentangling the RI channels for decision making, we are able to estimate the strength and prevalence of RI influences on investment across subjects: Roughly half of our subjects would reduce their allocation to a lucrative risky investment by an average of nearly 50% if that investment were linked to negative social externalities. The impact of positive externalities on investment decisions pales by comparison. We also find that subjects are significantly more pessimistic about investment outcomes when the investment is linked to a negative externality, even when controlling for the objective prospects of investments, while positive externalities do not appear to influence beliefs. The type, magnitude, and pervasiveness of the effects we identify can be readily incorporated into the modeling, and therefore equilibrium consequences, of RI preferences.

In magnitude, the asymmetries we find rival the strength of loss aversion estimates in the literature. Interestingly, evidence of charitable-giving motives appears to be largely absent in the data. The effect we find is not only high in magnitude, but also pervasive among a large portion of the subject pool. This suggests that, based on existing theoretical work, when extrapolated to a market setting our findings would translate into observable price impacts, but only for firms associated with negative externalities. Indeed, a survey of the empirical RI literature suggests not only a prevalence of negative screening – an important component of the majority of RI strategies – but that only "sin stocks" consistently exhibit greater discounting than warranted based on conventional risk adjustment (evidence for the overpricing of "angel stocks" tends to be mixed). Another of our novel experimental findings is that social preferences affect investors' subjective probabilities about their investments. Although this latter effect is modest, it reflects the importance that social preferences can have on how investors process information (e.g., update their beliefs).

Responsible investing has become an increasingly important aspect of individuals' investment opportunity sets. Theory and empirical evidence have demonstrated that growing tastes for responsible investing can impact asset pricing. Our findings help refine existing facts and insights by pointing to novel drivers of responsible investment. Importantly, our results have implications for how to think about incorporating social preferences into theoretical models because they demonstrate not only how preferences and beliefs are affected but also the large heterogeneity that exists in these effects. In addition, our results have implications for policy. In particular, the strong asymmetric effects we find suggest that, from an investor's perspective, the marginal benefit of reducing harm is much greater than the marginal benefit of doing good.

References

Abdellaoui, M., Baillon, A., Placido, L., and Wakker, P. (2011): The rich domain of uncertainty: Source functions and their experimental implementation, *American Economic Review* 101, 695-723.

Abdellaoui, M., Bleichrodt, H. and Paraschiv, C. (2007): Loss aversion under prospect theory: A parameter-free measurement, *Management Science* 53(10), 1659-1674.

Bansal, R., Wu, D., and Yaron, A. (2022): Socially responsible investing in good and bad times, *Review of Financial Studies* 35, 2067-2099.

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

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

Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics: Applications and Foundations 1*, 2, 69-186.

Bialkowski, J., and Starks, L.T. (2018): SRI funds: Investor demand, exogenous shocks and ESG profiles, Unpublished working paper.

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

Bonnefon, J-F., Landier, A., Sastry, P. and Thesmar, D. (2022): The moral preferences of investors: Experimental evidence, *NBER working paper 29647*.

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

Brodback, D., Guenster, N. and Pouget, S. (2020): The valuation of corporate social responsibility: A willingness-to-pay experiment, Unpublished working paper.

Chava, S. (2014): Environmental externalities and cost of capital, *Management Science* 60, 2223-2247.

Chew, S.H., and Li, K.K. (2021): The moral investor: Sin stock aversion and virtue stock affinity, *Available at SSRN 3773971*.

Chowdhry, B., Davies, S.W. and Waters, B. (2019): Investing for impact, *Review of Financial Studies* 32, 864-904.

Cooper, D. J., and Kagel, J. (2016): Other-regarding preferences, *The Handbook of Experimental Economics* 2, 217-289.

Edmans, A. (2011): Does the stock market fully value intangibles? Employee satisfaction and equity prices, *Journal of Financial Economics* 101, 621-640.

Edmans, A., and Kacperczyk, M. (2022): Sustainable finance, *Review of Finance* 26, 1309-1313.

Engel, C. (2011): Dictator games: A meta study, *Experimental economics* 14, 583-610.

Fama, E., and French, K. (2007): Disagreement, tastes and asset prices, *Journal of Financial Economics* 83 667-689.

Flammer, C. (2021): Corporate green bonds, *Journal of Financial Economics* 142, 499-516.

Geczy, C. C., Jeffers, J.S., Musto, D.K. and Tucker, A.M. (2021): Contracts with (social) benefits: The Implementation of impact investing, *Journal of Financial Economics* 142, 697-718.

Goldstein, I., Kopytov, A., Shen, L. and Xiang, H. (2021): On ESG investing: Heterogeneous preferences, information and asset prices, Unpublished working paper.

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

Heeb, F., Kölbel, J., Paetzold, F. and Zeisberger, S. (2021): Do investors care about impact? Unpublished working paper.

Heinkel, R., Kraus, A., and Zechner, J. (2001): The effect of green investment on corporate Behavior, *Journal of Financial and Quantitative Analysis* 36, 431-449.

Hong, H. and Kacperczyk, M. (2009): The price of sin: The effects of social norms on markets, *Journal of Financial Economics* 93, 15-36.

Humphrey, J. E., Lee, D.D,, and Shen, Y. (2012): Does it cost to be sustainable? *Journal of Corporate Finance* 18, 626-639.

Hwang, C.Y., Titman, S., and Wang, Y. (2022): Investor tastes, corporate behavior, and stock returns: An analysis of corporate social responsibility, *Management Science* 68, 7065-7791.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of political Economy*, 98(6), 1325-1348.

Kahneman, D., and Tversky, A. (1979). Prospect theory: An analysis of decision under risk, *Econometrica,* 47, 263-291.

Kensinger, E. A., & Schacter, D. L. (2008). Memory and emotion. *Handbook of emotions*, 3, 601-617.

Kuhnen, C. (2015): Asymmetric learning from financial information, *Journal of Finance* 70, 2029-2062.

Krueger, P. (2015): Corporate goodness and shareholder wealth, *Journal of Financial Economics* 115, 304-329.

Krueger, P., Sautner, Z. and Starks, L.T. 2020, The importance of climate risk for institutional investors, *Review of Financial Studies* 33, 1067-1111.

Larcker, D.F. and Watts, E.M. (2020): Where's the greenium? *Journal of Accounting and Economics* 69, Article 101312.

Liang, H., Sun, L., and Teo, M. (2022): Responsible hedge funds, *Review of Finance* 26, 1585-1633.

List, J. (2007): On the interpretation of giving in dictator games, *Journal of Political Economy* 115), 482-493.

Luo, H.A. and Balvers, R. (2017): Social screens and systematic investor boycott risk, *Journal of Financial and Quantitative Analysis* 52, 365-399.

Mellers, B., Schwartz, A., & Ritov, I. (1999). Emotion-based choice. *Journal of Experimental Psychology: General*, 128(3), 332.

Möbius, Markus M., Niederle, M., Niehaus, P. and Rosenblat, T.S. (2022): Managing self-confidence: Theory and experimental evidence, *Management Science* 68:11, 7793-7817

Oehmke, M., and Opp, M. (2020): A theory of socially responsible investment, Unpublished working paper.

Pastor, L., Stambaugh, R. and Taylor, L. (2021): Sustainable investing in equilibrium, *Journal of Financial Economics* 142, 550-571.

Pedersen, L.H., Fitzgibbons, S. and Pomorski, L. (2021): Responsible investing: The ESG-efficient frontier, *Journal of Financial Economics* 142, 572-597.

Prelec, D. (1998): The probability weighting function, *Econometrica* 66, 497-528.

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

Slovic, P., and Lichtenstein, S. (1971): Comparison of Bayesian and regression approaches to the study of information processing in judgment, *Organizational behavior and human performance*, *6*(6), 649-744.

Starks, L., (2023): Presidential address: Sustainable finance and ESG issues – *value* versus *values*, *Journal of Finance* 78, 1837-1872.

Statman, M., and Glushkov, D. (2009): The wages of social responsibility, *Financial Analysts Journal* 65, 33-46.

Stroebel, J., and Wurgler, J. (2021): What do you think about climate finance? *Journal of Financial Economics*, forthcoming.

Svenson, O. (1981): Are we all less risky and more skillful than our fellow drivers? *Acta psychologica*, *47*(2), 143-148.

Tversky, A. (1973): On the psychology of prediction, *Psychological Review*, *80*(4), 237-251.

Tversky, A., and Kahneman, D. (1992): Advances in prospect theory: Cumulative representation of uncertainty, *Journal of Risk and Uncertainty* 5, 297-323.

US SIF (2020): Report on US sustainable, responsible and impact investing trends 2020, Washington, DC: US SIF.

Yoo, P. (2022): ESG investing: A tale of two preferences, *SSRN Working Paper Series*, https://ssrn.com/abstract=4069015.

Zerbib, O. D. (2022): A sustainable capital asset pricing model (S-CAPM): Evidence from environmental integration and sin stock exclusion, *Review of Finance* 26, 1345-1388.

Appendix A: Experiment Snapshots

Panel I: Initial round allocation screen

First Block, Trial One

You are about to begin a new trial consisting of six investing rounds.

The **computer will now randomly select the type of** stock you will be facing during this trial.
There is a **50% chance that the computer will select a high payoff stock** and a 50% that it will select a low payoff stock.

A **high payoff doubles with probability 2/3** and halves with probability 1/3.
A **low payoff stock doubles with probability 1/3** and halves with probability 2/3.

Estimate the probability that this is the **high payoff** stock.

Please enter a number between 0 and 100:

You have **100 ECU**. Enter the amount you wish to invest in the stock in the next round. The rest will be invested in risk-free cash.

Stock:

Panel II: Outcome screen after three rounds

Trial: 1

The stock outcome for Round 3 is: Halve (↓)

History of Decisions and Outcomes

Round Number	Probability Estimate Prior to Outcome	Stock Allocation Prior to Outcome	Outcome	Winnings from Round
1	20	50	Halve (↓)	75
2	30	50	Double (↑)	150
3	40	40	Halve (↓)	80
4				
5				
6				

Recall that **high payoff** stocks will double with a probability of 2/3 and halve with a probability of 1/3.

Low payoff stocks will double with a probability of 1/3 and halve with a probability of 2/3.

Estimate the probability that this is the **high payoff** stock.

Please enter a number between 0 and 100:

You have **100 ECU**. Enter the amount you wish to invest in the stock in the next round. The rest will be invested in risk-free cash.

Stock:

Panel III: Social issues

What Social Issues Are Important to You?

Below are a number of current corporate social issues, listed in alphabetical order. Please rank these issues from 1 to 6 in terms of how much you care about each issue. A ranking of 1 means you care the most and a ranking of 6 means you care the least. You can rank by clicking and dragging. If you would like to keep the current ranking simply click and slightly drag any of the options.

Gender discrimination (example: not hiring women for jobs for which they are qualified)

Human Trafficking (example: forcing children to work as slave labour in factories or on farms)

Refugees (example: trading with dictatorial regimes which have resulted in mass fleeing from the country)

Poverty (example: underpaying workers in developing countries)

Animal welfare (example: infecting laboratory monkeys with diseases to test pharmaceutical drugs)

Environment (example: releasing carbon dioxide into the atmosphere)

What Social Issues Are Important to You?

You selected environment as your top social issue. Below is a description of two environmental non-profits.

The Rainforest Alliance works to conserve biodiversity and ensure sustainable livelihoods by transforming land-use practices, business practices and consumer behaviour.

Conservation Strategy Fund sustains natural ecosystems and human communities through strategies powered by conservation economics. Our trainings, analyses and timely expertise make development smarter, quantify the benefits of nature, and create enduring incentives for conservation.

Please select a non-profit organization to link to your trading profits:

- The Rainforest Alliance
- The Conservation Strategy Fund

Panel IV: Positive block instruction screen and outcome screen

The Conservation Strategy Fund gains by your stock investment

You are about to begin a block of two trials. Each trial consists of six investment rounds. You will be asked to make the same decision as in the practice trial.

At the beginning of each of the two trials the computer will randomly choose whether the stock for that trial is a high or low payoff stock. You will then face the series of six investment rounds.

In each investment round you will be asked how much you wish to allocate to the stock and to cash. After the stock outcome is determined, you will be asked to indicate the probability that you are facing a high payoff stock and how much you trust your probability estimate.

One round may be randomly chosen as your actual payoff round. If that round is chosen as your payoff round, in addition to paying you, **the researchers will DONATE the same amount of money as your payoff from your stock investment to your designated non-profit**. For example, if you earn 135 ECU from investing including 70 ECU from your stock investment, your chosen non-profit will receive 70 ECU.

The payment to the non-profit will come out of the researchers' funding, not from your trading payoff. You will be sent an email documenting our payments to the various non-profit organizations involved in this experiment in a few days.

Your designated non-profit for this block of three trials is: The Conservation Strategy Fund

Non-profit gains by your stock investment

Trial: 1

The stock outcome for Round 3 is: Double (↑)

History of Decisions and Outcomes

Round Number	Probability Estimate Prior to Outcome	Stock Allocation Prior to Outcome	Outcome	Winnings from Round (if round is selected)	Payment to The Conservation Strategy Fund (if round is selected)
1	40	60	Double (↑)	160	120
2	60	45	Halve (↓)	77.5	22.5
3	50	50	Double (↑)	150	100
4					
5					
6					

Recall that **high payoff** stocks will double with a probability of 2/3 and halve with a probability of 1/3.

Low payoff stocks will double with a probability of 1/3 and halve with a probability of 2/3.

Estimate the probability that this is the **high payoff** stock.

Please enter a number between 0 and 100:

You have **100 ECU**. Enter the amount you wish to invest in the stock in the next round. The rest will be invested in risk-free cash.

Stock:

Panel V: Negative block instruction screen and outcome screen

Block: SPCA International loses by your stock investment

You are about to begin a block of two trials. Each trial consists of six investment rounds. You will be asked to make the same decision as in the practice trial.

At the beginning of each of the two trials the computer will randomly choose whether the stock for that trial is a high or low payoff stock. You will then face the series of six investment rounds.

In each investment round you will be asked how much you wish to allocate to the stock and to cash. After the stock outcome is determined, you will be asked to indicate the probability that you are facing a high payoff stock and how much you trust your probability estimate.

One round may be randomly chosen as your actual payoff round. If that round is chosen as your payoff round, in addition to paying you, **the researchers will DEDUCT from funds allocated to the non-profit the same amount of money as your payoff from your stock investment**. For example, if you earn 135 ECU from investing including 70 ECU from your stock investment, your chosen non-profit will have 70 ECU deducted from their balance.

The deduction from the non-profit will have no effect on your trading payoff. You will be sent an email documenting our payments to the various non-profit organizations involved in this experiment in a few days.

Your designated non-profit for this block of three trials is: SPCA International

Block: Non-profit loses by your stock investment

Trial: 1

The stock outcome for Round 3 is: Double (↑)

History of Decisions and Outcomes

Round Number	Probability Estimate Prior to Outcome	Stock Allocation Prior to Outcome	Outcome	Winnings from Round (if this round is selected	Deduction from SPCA International (if this round is selected)
1	50	30	Halve (↓)	85	-15
2	20	40	Halve (↓)	80	-20
3	15	30	Double (↑)	130	-60
4					
5					
6					

Recall that **high payoff** stocks will double with a probability of 2/3 and halve with a probability of 1/3.

Low payoff stocks will double with a probability of 1/3 and halve with a probability of 2/3.

Estimate the probability that this is the **high payoff** stock.

Please enter a number between 0 and 100: ______

You have **100 ECU**. Enter the amount you wish to invest in the stock in the next round. The rest will be invested in risk-free cash.

Stock: ______

Appendix B: Survey to establish external validity

Below is an example of a survey question eliciting subjects' preference for "reducing good" versus "causing harm". The full survey is available at https://uqbel.az1.qualtrics.com/jfe/form/SV_3VID8WMyu4q05z8.

A local town council debates the use of municipal funds to house and support people experiencing homelessness. Because of the limited funds available for all of the city needs, the council is evenly split on the measure, and the final vote resides with the elected mayor who is undecided (but will have to decide one way or another in a week). It is impossible to know which way the Mayor currently leans.

If you had to make a choice between the following alternatives, which would you choose

- **Donate** $1,000 to a firm that will lobby the Mayor to vote **against** helping people experiencing homelessness.
- **Cut** $1,000 of funding from a charity that will lobby the Mayor to vote **for** helping people experiencing homelessness.
- The choices are equally bad.

Appendix C: Closer Examination of Subject Heterogeneity

In a naïve approach we would regress allocations on subjective probabilities for each subject and treatment, but this presents two major challenges. First, a subject-by-subject regression analysis lacks power because we have only twelve decisions for each subject per treatment. Second, subject responses reflect experimental error for a variety of reasons.[28] Because allocations are bounded between 0 and 100 ECU, experimental noise can bias observations away from the bounds and correspondingly bias inferences about individual allocation responses to perceived probabilities.[29] To address these issues, we adopt a reduced form model relating noisy optimal allocations to subjective probabilities. We then estimate individual-level parameters as random effects. This allows us to efficiently quantify individual preferences as well as attempt to control for edge biases arising from noise.

We begin by noting that most utility models predict an optimal allocation that roughly resembles a sigmoid function of subjective probability.[30] Consistent with that, we assume that noisy individual allocations, in a given treatment, can be described in reduced form as follows:

$$ECU_i^T = a_i + b_i f(Prob_i) + e_i, \qquad (C1)$$

where the transformed allocation, $ECU^T \in \Re$, is unbounded and related to the allocation choice via the sigmoid transformation,

$$ECU = 100/(1 + \exp(-ECU^T)). \qquad (C2)$$

In Equation (C1), i refers to the subject, $f()$ is an increasing function to be specified soon, a_i and b_i are subject-specific constants, and e_i is experimental noise. A standard sigmoid

[28] Subjects make mistakes, get confused, or simply find it difficult to maintain the same level of engagement throughout the rounds of an experiment. There is also the possibility that they enjoy injecting a random component to their choices. Holding constant the subjective probability and treatment, the average filtered subject exhibited a standard deviation of 11 ECU in their allocations.

[29] While this bias can also impact the estimates in Table II, within-subject estimates are more sensitive to the noise-induced bias. Indeed, consistent with a bias-driven deviation from the regression model, the residuals in Regression (2) of Table II exhibit a U-shape magnitude with respect to the perceived probability (roughly highest at the boundaries).

[30] A sigmoid function takes the form, $y = A\,(1 + \exp(a - bx))^{-1}$ with $b \geq 0$.

corresponds to the case where *f* is linear. Because ECU is convex in ECU^T for low values and concave for high values, this formulation models how experimental noise biases allocations away from the boundaries of [0,100].[31] The choice of *f*() reflects behavior in the absence of noise. For instance, if *f*(*p*) is finite at *p=0*, then the decision maker would be risk-loving because the allocation to the stock (an actuarially fair gamble at *p*=0) would be greater than zero. Because b is a measure of the sensitivity of allocation to a first-degree stochastically dominating shift in the payoff distribution of the stock, one can interpret *b*>0 as a measure of local risk tolerance with higher *b* signifying higher risk tolerance. We model

$$f(p) = \ln\left(\frac{p}{1-p}\right). \qquad (C3)$$

Our modeling choice implies global risk aversion and that, absent noise, as subjective probability approaches one the allocation will tend to the full endowment of 100 ECU. This lends parsimony to modeling the dependence of allocation on subjective probability.[32]

To address the issue of statistical power, we adopt a random effects framework estimated to allow for treatment differences. That is, we allow each subject to have a different average level of investment (intercept, a_i) and a different allocation sensitivity to probabilities (slope, b_i) in each of the treatments. In a random effects framework, the subject-level coefficients are assumed to be drawn from a distribution whose mean and standard deviation are estimated. A benefit of this approach over individual regressions is the joint estimation of covariance of distinct random effects. The estimated model is

$$ECU^T_{\tau,i,n} = \ a_{\tau,i} + b_{\tau,i} f\left(Prob_{\tau,i,n}\right) + e_{\tau,i,n}. \qquad (C4)$$

[31] This is an implication of Jensen's Inequality.

[32] To avoid infinities in expressions (2) and (3), extreme allocations of 0 or 100 ECU are adjusted to 0.1 or 99.9 ECU, respectively, and extreme reported probabilities of 0 or 1 are adjusted to 0.001 or 0.999. The qualitative conclusions of our analysis are robust to simply excluding these observations. They are also robust to alternative specifications such as, $f(p) = p$, or $f(p) = ln(p)$, or $f(p) = c\,ln(p) + dp$, or $f(p) = c\,ln(p) + \ ln(1-p)$.

where τ denotes the treatment, i, the subject ID (corresponding to the random effect), n, the round number for the given treatment, and $Prob_{\tau,i,n}$, the probability reported by the subject in round n of the given treatment.

Results are reported in Table C-I and correspond to the estimated means, standard deviations, and correlations of the distribution of random coefficients.[33] Table C-I also reports the residual standard deviation. Consistent with our prior results, we observe that in the Negative treatment, relative to the Neutral one, both the baseline allocation, mean(a), and the sensitivity to probabilities, mean(b), are significantly lower. By contrast, the two Positive treatment mean coefficients do not allow for a simple ranking relative to the Neutral treatment estimates (the joint difference is statistically insignificant with a p value of 36%).

Table C-I: Random effects regression of transformed allocation on reported probabilities

The table reports random-effect regressions of transformed allocations on reported probabilities estimated as specified in Equation (C4). a refers to intercept and b to sensitivity to probabilities. The error term is assumed to be i.i.d. across subjects and rounds, within treatment. Standard errors are in parentheses.

Treatment	mean(b)	sd(b)	mean(a)	sd(a)
Negative	0.652***	0.699**	-1.733***	1.880***
	(0.082)	(0.081)	(0.177)	(0.132)
Neutral	0.966***	0.581***	-0.751***	1.651***
	(0.069)	(0.073)	(0.156)	(0.117)
Positive	0.903***	0.857***	-0.583***	1.690***
	(0.095)	(0.087)	(0.159)	(0.119)
	sd(e_{Neg})	sd(e_{Neu})	sd($ePos$)	N
	1.904***	1.805***	1.684***	4500
	(0.038)	0.036	0.034	

Note: *p<0.1; **p<0.05; ***p<0.01

Second, the table shows that the slope dispersion across subjects, sd(b), is estimated to be larger in the Negative and Positive treatments, compared with the Neutral treatment. This

[33] We only report significant random effect correlations.

suggests pronounced heterogeneity in treatment effects across subjects. Finally, we observe large and heterogeneous levels of residual variation across the treatments. The residual magnitudes justify our effort to adjust for edge-induced biases.[34]

Figure C-1, Panel A, plots the estimated allocation using the mean coefficients from Table C-I and setting the experimental noise to zero. The plot includes 95% confidence interval bars for the Positive and Negative treatments.[35] From this figure we observe that the negative treatment effect is most pronounced when the subjective probability is higher than 50%. The figure is consistent with the absence of an average charitable giving motive which, if present, would imply a significantly higher allocation in the Positive treatment at low probabilities. As with Figure 1, the Positive treatment allocations are generally above those of the Neutral treatment, but we stress that the two Neutral and Positive treatment curves cannot be differentiated statistically because they are generated by coefficients that are (jointly) not statistically different.

Figure C-1, Panel B, plots the magnitude of the deviations of each of the two treatments from the Neutral case. Holding constant the subjective probability, the magnitude of the externality is the same in the Negative and Positive treatment (with the opposite sign, of course). The plotted deviations from the Neutral treatment can therefore be viewed as responses to an equal-magnitude positive versus negative stimuli. From the plot, when the subjective probability is at 50%, the estimated model response is more than four times larger to a Negative "stimulus" than a positive one. This is consistent with the previous estimates and rivals loss-aversion in strength.

[34] The random effects model estimates correlations between the *a*'s and *b*'s, some of which are found to be statistically different from zero. In the interest of brevity, we omit these results.

[35] The Neutral treatment confidence intervals, like the Neutral treatment point estimates, are close to those of the Positive treatment. We omit these for visual clarity.

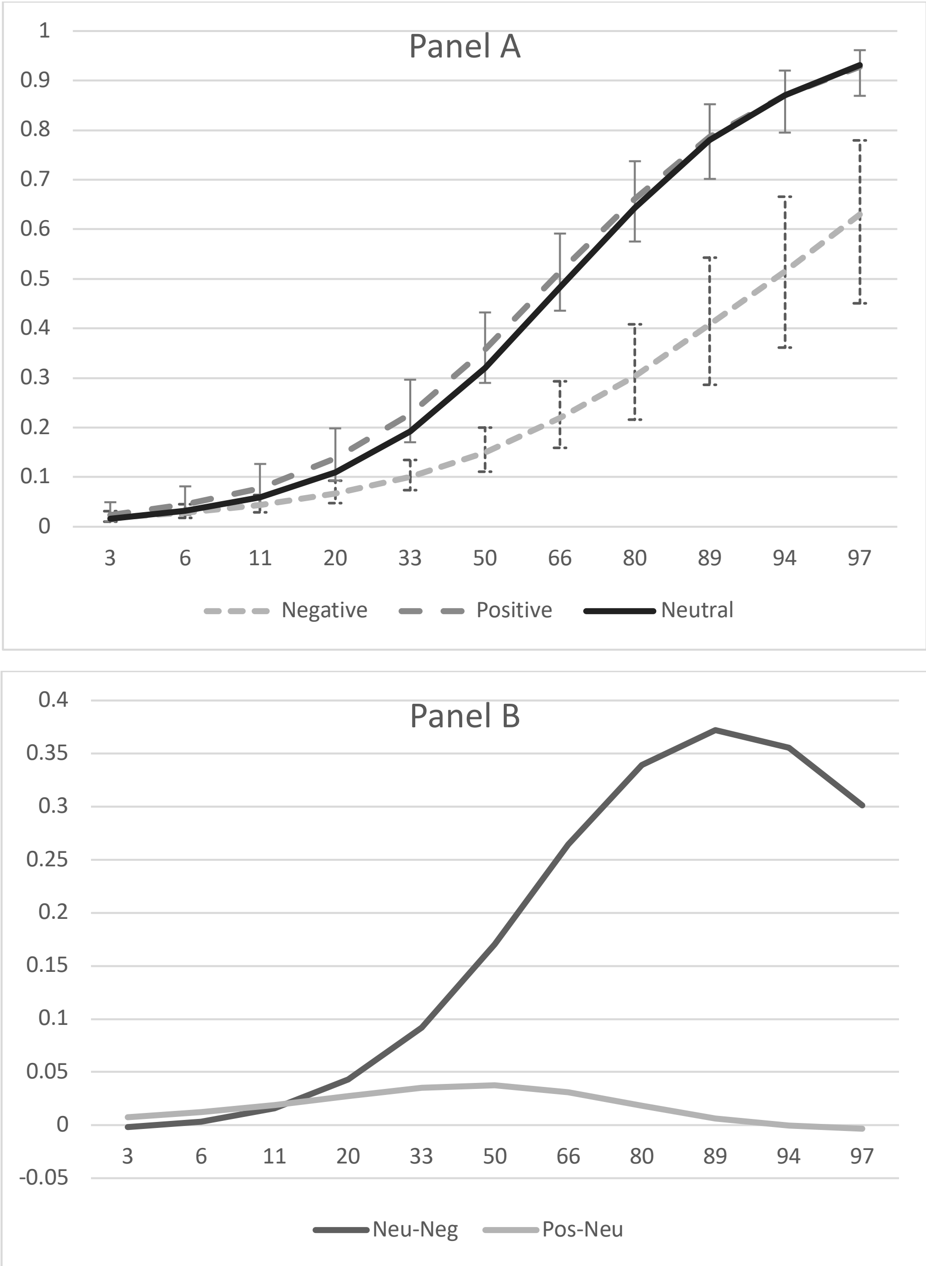


Figure C-1: Expected allocations across treatments
This figure depicts estimated model predictions of the treatment effects on allocation (y-axis) against subjective probability that the stock is of the high type (x-axis). The estimated model prediction sets experimental noise to zero. By contrast, Figure 1 is based on aggregated statistics that potentially include bias from experimental noise. Panel A reports the expected allocation across treatments while Panel B reports the expected allocations relative to the neutral treatment (i.e., neutral-negative and positive-neutral).

To compare model predictions to the simple statistics from Table I, Panel A, we use the model estimates of each subject's intercept and slope parameters to calculate that subject's *expected allocations* in the treatment without experimental noise. We do this using the theoretical distribution of objective stock probabilities in the six rounds.[36]

Table C-II reports the average over subjects of their predicted expected allocation in each of the treatments. Consistent with the results of Table I, the expected allocations are lower in the Negative treatment (0.262) relative to the Neutral treatment (0.373) which, in turn, is not far from the expected allocation in the Positive treatment (0.403). The magnitude of the treatment effect is similar to that documented in Table I and corresponds to a reduction of nearly 30% in allocations in going from the Neutral to the Negative treatment.

Table C-II: Average of expected allocations in each treatment

The table reports the expected allocation averaged across subjects in each of the treatments based on subject-level coefficient estimates of the mixed regression in Equation (C4) and assuming no experimental noise (the residual is set to zero).

Treatment	Obs	Mean	Std. Dev.	Min	Max
Negative	125	0.262	0.183	0.001	0.783
Neutral	125	0.373	0.205	0.015	0.996
Positive	125	0.403	0.206	0.021	0.998

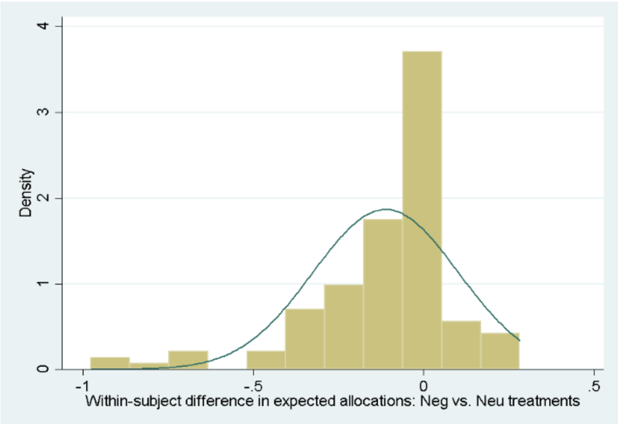
To examine the heterogeneity of the treatment effect across subjects, we compute for each subject the difference in expected allocation between the Negative and the Neutral treatments, and again the difference in expected allocation between the Positive and the Neutral treatments. This analysis allows us to compute two relative treatment effect measures per subject. We plot the distribution of these relative treatment effects in Figure C-2. Under the null that differences in behavior across treatments are noise, the distribution of relative

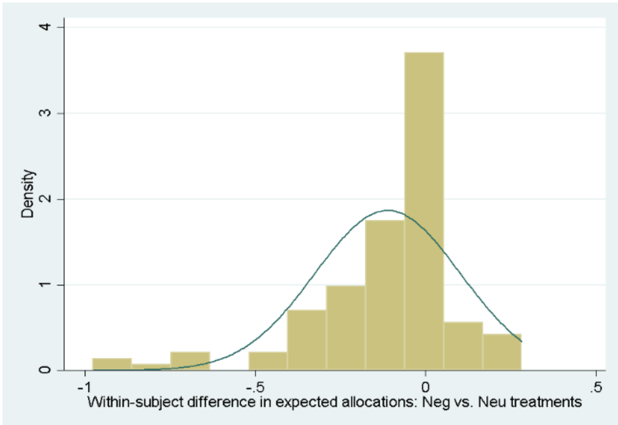
[36] For example, using the binomial distribution, we calculate that the stock should objectively be deemed to have a 50% chance of being the high-paying type in 29% of the rounds: This is sure to be the case in all of the first rounds, in 4 of 9 instances of the third round, and in 24/81 instances of the fifth round (and in none of the second, fourth, and sixth rounds). Adding up and dividing by six possible rounds we get (1+4/9+24/81)/6 = 0.29.

treatment effects should be symmetric about zero. The histograms in Figure 3 suggest that this roughly holds for the Positive treatment but that the same is not true for the Negative treatment.

Figure C-2: Distribution of allocations -- Histograms

The figure reports the distribution of subjects' expected allocations in the two treatments relative to their expected allocation in the Neutral treatment. Expectations are calculated using subject-level coefficient estimates of the mixed regression in Equation (C4) and assuming no experimental noise (the residual is set to zero).





To quantify treatment-dependent heterogeneity across subjects, we estimate a 2-component finite mixture model of normal distributions to each of the histograms. The results, reported in Table C-III, suggest that the vast majority of subjects (just over 90%) are drawn from a population that expects to allocate only two more ECUs in the Positive treatment (component 1) than in the Neutral treatment, while the remaining part of the population may allocate much more (14 ECUs) but their mean allocation is not statistically distinguishable from zero (component 2). By contrast, roughly half the subjects are drawn from a population that expects to allocate 20 ECUs *less* in the Negative treatment (relative to the Neutral treatment), while the remaining expect to allocate 3 ECU less. This analysis demonstrates that the treatment effect is not just strong in aggregate but also pervasive across subjects. The finding that roughly half the subjects are highly sensitive to the Negative treatment is consistent with the survey finding by Bauer et al. (2021) that two out of three Dutch pension plan participants were in favor of RI mandates.

Table C-III: Distribution of allocations: Finite-mixture model analysis

The table reports a decomposition of each histogram in Figure C-2 into a mixture of two normal distributions. Means, standard deviations, and mixture probability are estimated for the data depicted in each of the histograms.

Within-subject difference in expected allocations relative to the Neutral treatment		
	Negative treatment	Positive treatment
Mean of component 1	-0.0296**	0.0211**
	(0.009)	(0.007)
Mean of component 2	-0.195***	0.12
	(0.039)	(0.144)
SD of component 1	0.0500***	0.0687***
	(0.009)	(0.009)
SD of component 2	0.274***	0.404**
	(0.027)	(0.122)
Probability of component 1	0.505	0.916***
	(0.074)	(0.052)
N	125	125

Note: *p<0.1; **p<0.05; ***p<0.01

To contextualize these estimates consider that, based on the estimates from Table C-III, 40%-50% of subjects should exhibit Type 2 tendencies in the Negative treatment and Type 1 tendencies in the Positive treatment. This translates into an additional allocation of 2 ECUs in the Positive treatment and -20 ECU in the negative treatment. The asymmetry multiplier, in this case, is ten! The remaining subjects who exhibit Type 1 tendencies in the Positive treatment will exhibit a smaller asymmetry multiplier of 1.4 that is still significantly different from 1. These two populations, both of which exhibit significant asymmetric multipliers, account for 90% of the subject pool.

Finally, to obtain further confirmation that a Negative treatment effect is pervasive among participants, we plot the expected allocation in the Negative treatment against the Positive treatment, by subject, and relative to a 45-degree line (see Figure C-3). First, the figure shows that most subjects fall under the 45-degree line, consistent with allocations in the Negative treatment being lower than in the Positive treatment, even when accounting for differences in baseline levels of allocations. Second, the asymmetry in distribution relative to the 45-degree line is observed for virtually all levels of Positive treatment allocations. Thus, the sensitivity to negative externalities is found among all subjects – those that appear to care about positive externalities and those that do not.

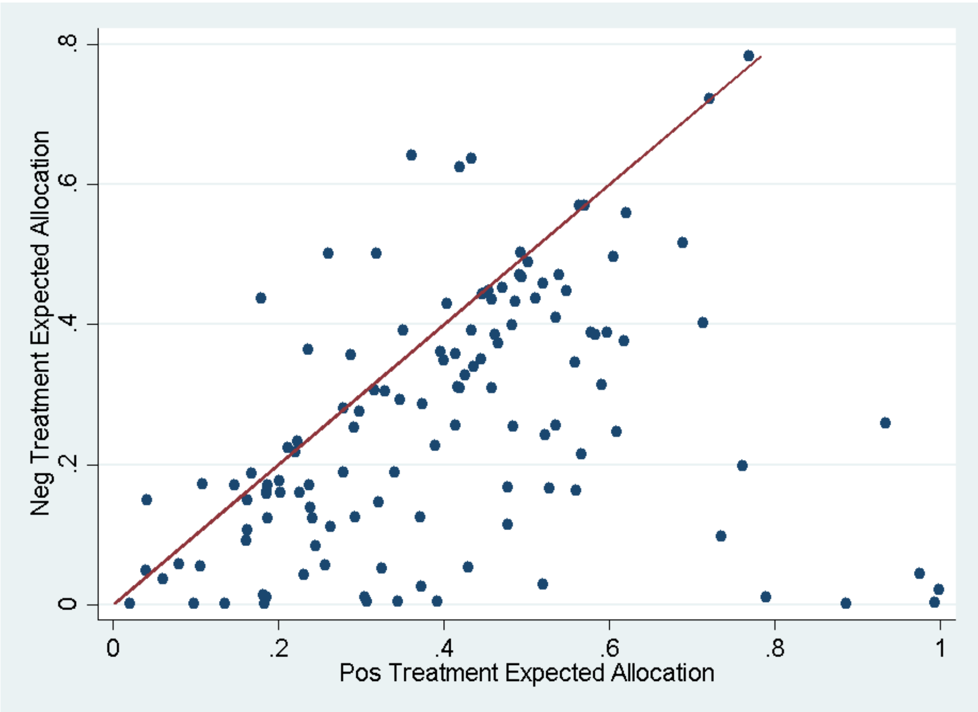


Figure C-3: Expected allocations in positive and negative treatments
The figure reports the expected allocations in the Positive and Negative treatments based on model estimates (see Table C-I). Each dot corresponds to a subject.

Appendix D: An Analysis of Learning

To further understand how considerations of externalities distort beliefs, in this analysis we examine how subjects learn in the experiment. Because probability estimates are submitted after observing realization of payoffs from the stock investment, we can investigate whether, and how, the treatments affect learning.

An extensive literature studies experimentally and theoretically how learning in various settings deviates systematically from the Bayesian framework (e.g., Tversky, 1973; Slovic and Lichtenstein, 1971; Svenson, 1981). We build on the work of Mobius et al. (2022) who nest two important deviations from Bayesian updating in a simple linear framework by transforming priors and posteriors into log odds. The nested deviations include asymmetric updating (responding differently to positive and negative signals) and conservatism (interpreting signals as less informative than they are). Specifically, we estimate the following linear regression:

$$logit(\mu_{i,t}) = \delta\, logit(\mu_{i,t-1}) + \beta_H\, I(S_{i,t} = H)\lambda_H + \beta_L\, I(S_{i,t} = L)\lambda_L + \varepsilon_{i,t} \qquad \text{(D1)}$$

Where $\mu_{i,t}$ is the reported probability by subject *i* in period *t*, $logit(x) = \ln(x/(1-x))$, $\lambda_H = \ln(2)$ is the log odds of the probability that the stock is of the "High" type, $\lambda_H = -\ln(2)$ is the log odds of the probability that the stock is of the "Low" type, $I(S_{i,t} = H)$ is an indicator function for a round in which the stock doubles, and $I(S_{i,t} = L)$ is an indicator function for a round in which the stock halves. It is straightforward to check that the Bayesian case corresponds to a setting where $\delta = \beta_H = \beta_L = 1$. Any difference between β_H and β_L implies an asymmetry in updating with respect to positive versus negative information about the stock's performance. Deviation of the coefficients from one corresponds to subjects' under- or overreaction to information.

We estimate this regression for each treatment separately while clustering standard errors by subject. The regression results are presented in Table D1, panel A. Consistent with Mobius et al. (2022), we find that subjects generally tend to underweight new information when updating, as compared with the Bayesian predictions (panels A and B). In the Neutral treatment, there does not appear to be asymmetry in underweighting positive versus negative new information about the stock. This contrasts with the results in Mobius et al. (2022), potentially because, aside from financial consequences, subjects are not emotionally linked to whether the information is positive or negative.[37] Consistent with the hypothesis that the Positive and Negative treatments influence subjects' emotional connection to stock outcomes, Panel C presents evidence of asymmetry in the Positive and (especially) Negative treatments. In both cases, positive information about the stock type is discounted relative to negative news. However, the asymmetry is much more pronounced in the Negative treatment.

When we test for treatment effects, comparing the Neutral treatment to both the Positive and Negative ones, we find interesting results. First, there does not appear to be a statistically significant difference between the Neutral and Positive treatments. Second, the only significant difference that we find between the Neutral and Negative treatments is observed with respect to the response to positive information. Namely, subjects respond less to positive information in the Negative treatment relative to the Neutral one. This aligns with our observations about the role of the Negative treatment in distorting perceptions of outcomes: The Positive and Neutral treatments are nearly indistinguishable, while the Negative and Neutral treatments differ significantly.

Although the results on the asymmetry in updating in the Negative treatment are intriguing, their economic magnitude is rather small. This is consistent with the overall pattern observed in our study where belief distortion plays a secondary role when compared to the treatment effects on preferences (holding beliefs constant). To quantify this, we use the estimated coefficients to compute the posterior probability of the average subject in the Negative relative to the Neutral treatments after observing a sequence of five positive signals

[37] In Mobius et al. (2022), the information conveys personal information in that it is about a subject's IQ.

or a sequence of five negative signals. We find that after a sequence of five positive signals, subjects in the Negative treatment are predicted to report a probability of 83.4%, relative to 86.7% in the Neutral one: A 3.3% decline in reported probability. The difference after a sequence of five negative signals is only 1.8% (10.6% relative to 12.4%).

D1: Learning across treatments

The table reports the regression results corresponding to Eq. (D1), in which sequential learning is allowed to deviate from the Bayesian null for both a base-rate neglect as well as asymmetry in response to negative and positive information. The model is estimated separately for each treatment cell (Panel A); test of coefficients' deviations from the Bayesian null are included as asterisks; differences in estimated coefficients across treatments are tested (Panel B).

Panel A: Subjective probabilities across treatments

	(1)	(2)	(3)
Variables	Neutral	Positive	Negative
δ	0.889**	0.905**	0.956
	(0.03)	(0.04)	(0.04)
β_H	0.674**	0.591**	0.508**
	(0.04)	(0.04)	(0.04)
β_L	0.704**	0.708**	0.672**
	(0.05)	(0.04)	(0.04)
Observations	1,159	1,169	1,125
R-squared	0.68	0.73	0.76

Panel B: Testing for treatment effect

(p-values)	(1)	(2)
Test	Neutral==Positive	Neutral==Negative
δ	0.6991	0.1734
β_H	0.0604	0.0006
β_L	0.9295	0.4391