When Do Open-Ended Answers Align with Choices? Evidence from Sustainable Investing

Tobias Wekhof*

November 3, 2026

Abstract

Financial advisers are increasingly required to ask clients about their sustainability preferences, often with a checklist. Ticking a box costs less effort than a written answer, but an unticked box is a clearer "no" than an empty text field. In a randomized survey, 3,059 Swiss retail investors described their investment priorities twice, in writing and on a checklist, and then chose between a hypothetical high-fee sustainable fund and a less sustainable low-fee fund. The value–action gap is the share choosing against a stated value. Because the two values compete in this choice, I measure each value's gap net of the other. Moving from written answers to the checklist raises the fee gap from 40% to 51% but lowers the sustainability gap from 16% to 10%. A structural model traces both effects mainly to the low salience of fees: many investors tick fees but do not write about them. Three other surveys from published papers, on inflation expectations and political identities, show the same two effects.

JEL Classification: G51; G11; C83; Q56

Keywords: value–action gap, household finance, open-ended questions, question format, survey methods, text analysis, natural language processing

*Geneva Finance Research Institute, University of Geneva, Geneva, Switzerland; ETH Zurich, Center of Economic Research, Zurich, Switzerland; email: tobias.wekhof@unige.ch.

I thank Anna Alberini, Ursa Bernadic, Massimo Filippini, Cloé Garnache, Katharina Holzheu, Sébastien Houde, Jon Krosnick, Markus Leippold, Aleksei Minabutdinov, Samuel Schenk, Kurt Sweat, Johannes Wohlfart, and the participants of the 2024 European Meeting of the Economic Science Association, the 18th International Behavioral Finance Conference, and the SSES Annual Congress 2025 for their valuable comments. I further thank Frank Wolak and the PESD team at Stanford University for their hospitality. The financial support from the 2° Investing Initiative (2DII) and Climate-KIC is gratefully acknowledged. This paper is part of a project that has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 870245. This study was registered in the AEA RCT Registry as AEARCTR-0017723 (registered after the data collection). A previous version of this paper is available at SSRN under the title "Explaining the Attitude–Behavior Gap for Sustainable Investors: Open vs. Closed-Ended Questions" (https://dx.doi.org/10.2139/ssrn.4838449).

1 Introduction

Surveys are commonly used to elicit values that are difficult to infer from observed choices alone. A basic design decision is whether respondents write what matters to them with an open-ended response or endorse items from a closed-ended list. Written answers can be coded into the same topics as the list items, so both formats show how often each topic is reported. The literature suggests that question format can change these frequencies and even reverse the ranking of topics (Schuman and Presser 1979, Schuman and Scott 1987). Yet, little is known about how this distinction impacts the alignment of a survey question's answers with subsequent choices, which I measure by the so-called value–action gap: it is defined as the rate at which a respondent states that an attribute matters, yet makes a choice that does not reflect that stated value. I show that writing can make an endorsement more selective, while a checklist can make a non-endorsement more informative; when two values compete, one effect will dominate.

This paper considers question format as a selection for latent value strength. Individuals differ in how strongly they value a product attribute, and expressing a value costs attention and effort. Closed-ended questions lower this cost and admit weakly held values; open-ended questions raise the effort and select stronger convictions. Because stronger convictions are more likely to guide choices, when each value is considered on its own, open-ended answers should align with choices at least as well as closed-ended ones. The size of this advantage depends on the attribute: when moderately held values already translate into choices, closed-ended answers align similarly well; when the closed-ended question adds many weakly held values, they pool strong and weak values and align less well. I use salience to describe how readily an attitude comes to mind without priming, that is, whether it is top of mind (cf. Gennaioli and Shleifer 2010).[2]

I test this mechanism in the context of retail investing in a survey of 3,059 Swiss private investors, who described their investment priorities twice, in writing and on a checklist, in randomized order, and then chose between a hypothetical high-fee sustainable fund and a less sustainable low-fee fund. Regulators in the EU require financial advisers under MiFID-II to elicit every client's sustainability priorities (European Commission 2021). In practice, elicitation is often implemented via closed-ended questions (i.e., ticking a box); this raised the concerns in the industry that stated sustainability priorities may be artificially over-reported due to the elicitation format and may not correspond to subsequent choices (e.g., Strachan and Gupta 2022).

[2]The term follows the attitude-strength literature, in which attitudes that are important to respondents are more likely to be mentioned spontaneously (Krosnick 1990) and accessibility is measured by how quickly an attitude comes to mind (Fazio and Williams 1986).

The survey answers and the fund choice depend on the same latent conviction. A value is reported once the conviction passes a threshold, which is lower for ticking a box than for writing. The stronger the conviction, the more the value weighs in the choice. The two values compete in this choice: the investor trades off the funds' attributes, a sustainable investment objective against a lower fee, as in a logit regression of the choice on these attributes, the classic revealed-preference setting, with each attribute weighted by the investor's conviction; most investors who report both values give up low fees.[3]

The gap of the value being evaluated, which I call the target value, can therefore be computed with or without the competing value. Taking one value at a time, the fee gap includes all investors who report fee concern, including those who also report sustainability and give up low fees in this trade-off. Net of the competing value, it includes only fee reporters who do not report sustainability, as a regression does when it adds sustainability as a second covariate. The same holds for the sustainability gap, with fees as the competing value.

This paper's results show that the effect of the question format on the alignment of stated values with choices depends on the attribute salience. Considering one value at a time, sustainability answers align with choices about equally in both formats: 17% of open-ended sustainability mentions and 16% of closed-ended sustainability endorsers choose the low-fee fund. Fees look different. Under closed-ended elicitation, 63% of investors who endorse low-fee importance still choose the high-fee fund; under open-ended elicitation, 48% do. Open-ended elicitation selects a smaller, more behaviorally consistent group: a) fewer investors spontaneously express fee concern and b) those who do are more likely to act according to their stated values (i.e., choose the low-fee fund).

Net of the competing value, switching from an open to a closed-ended question changes the target value's gap in two ways: target broadening and competing-value screening. First, as one value at a time, the closed-ended question adds respondents who tick the target value without writing about it. Their convictions are weaker, so they act on it less often than writers do, which raises the gap. Second, an unticked box is a clearer "no" than an empty text field. Respondents who tick the competing value without writing about it now count as reporting it and drop out of the comparison group; since they choose against the target value more often than those who remain, the gap falls. One value at a time, no one is dropped, so this second effect cannot appear.

Net of the competing value, moving from open to closed-ended questions raises the fee gap from 39.7% to 50.5% but lowers the sustainability gap, which barely moved one value at a time, from 15.7% to 9.9%. The gaps move in opposite directions mainly because

[3]In their first answers, 78 percent of the investors who report both values choose the sustainable fund: 82 percent when the funds carry a sustainability rating and 65 percent when they do not.

of the low salience of fees: many investors tick fees without writing about them, but few do so for sustainability. With fees as the target, the closed-ended question therefore adds many weak fee endorsers but drops few respondents who selected sustainability from the comparison group, and the gap rises. With sustainability as the target, it adds few respondents who would have selected sustainability but drops many fee endorsers, and the gap falls.

To rationalize these patterns, I derive a structural model, conceptually following Akerlof and Kranton (2000). The model shows how likely an individual is to make a choice that aligns with their latent values about the choice attributes; then, it characterizes this likelihood of action as a function of relative value strength (measured by an individual's quantile in the latent distribution). It formalizes the link as a hybrid choice model (Ben-Akiva et al. 2002): the survey answers are thresholded readings of a latent conviction, and the fund choice is a logit in which each value's conviction raises the utility of the fund it favors. Each part follows existing models.

The structural model adds two dimensions to these comparisons: first, it distinguishes whether a value is reported from how much it weighs in the choice. Second, it separates the two margins, target broadening and competing-value screening, which the raw data can do only with second answers that the first question may influence. For fees, target broadening raises the gap more than the stricter sustainability screen lowers it; for sustainability, the stricter fee screen lowers the gap more than target broadening raises it, although this split is less certain.

In three other surveys from the literature, on inflation expectations (Andre et al. 2022) and on political identities (Bornschier et al. 2024, 2021), the same respondents answer an open and a closed question and give a forecast or a vote. In each example, the estimates show the same pattern: net of the other value, the closed-ended question raises the gap of the less salient value and lowers the gap of the more salient value.

This paper contributes to the debate on open- versus closed-ended elicitation by offering (i) evidence that question types differ in how well their answers align with subsequent choices, with the size and direction of this difference depending on the salience of the target and the competing attribute, (ii) a mechanism, modeled with standard parts, in which open- and closed-ended answers are two thresholds on one latent conviction, so that a change of question adds and removes identifiable groups of respondents. As a practical rule for elicitation: open-ended answers are selective positive signals, closed-ended non-selections are selective negative signals, and the format should therefore be chosen with both the target value and its competitor in mind.

The remainder of the paper proceeds as follows. Section 2 reviews related literature. Section 3 describes the survey and experimental design, as well as the other surveys, and

Section 4 presents descriptive evidence on open- versus closed ended responses. Section 5 displays the reduced form results. Section 6 develops the underlying microeconomic model, Section 7 introduces the structural estimation results and the evidence from other surveys, and Section 8 concludes.

2 Related Literature

This paper contributes to the literature on how survey question format shapes the measurement of stated values and their alignment with choices. It connects three bodies of work: (i) the open- versus closed-ended elicitation debate in survey methodology, (ii) applications of the value-action gap that evaluate stated values against subsequent decisions, with a particular focus on sustainable investing, and (iii) behavioral choice models that provide a micro-foundation for why survey questions can select on value strength.

2.1 Open- versus closed-ended elicitation

The methodological debate over open- versus closed-ended questions dates back to at least the 1940s (Krosnick 1999, Lazarsfeld 1944). Closed-ended formats are widely used because they are inexpensive to administer and easy to code, but they can shape answers through recognition and priming (Ferrario and Stantcheva 2022, Schuman and Presser 1979) or social desirability (Tourangeau et al. 2000).

The open- versus closed-ended literature shows that question form can change response frequencies, content, and topic rankings. Open-ended questions require respondents to retrieve and articulate considerations, making them more demanding and potentially filtering out low-salience considerations (Schuman and Presser 1979); and Reja et al. (2003) documents differences in response content and burden. Most importantly, Schuman and Presser (1979) and Schuman and Scott (1987) show that question format can reverse topic rankings; which is also confirmed by Wekhof and Houde (2023); other comparisons find largely stable orderings of the leading topics (Reja et al. 2003, Schuman et al. 1986).

Open-ended answers may also be more selective, containing mainly the concerns respondents consider important. Schuman et al. (1986) and Geer (1991) test whether open responses merely reflect concerns that are conspicuous, such as issues recently in the news, rather than important; both studies conclude that open responses reflect what respondents consider important. Schuman et al. (1986) also find that the less often a problem is mentioned in the open format, the larger its relative increase when it is offered as a closed alternative, which they attribute to the list bringing to mind problems that respondents would not have named spontaneously. Krosnick (1999) emphasizes satisficing (see also Krosnick 1991): when an optimal answer takes substantial effort, respondents

may settle for a merely satisfactory one, for example by searching memory less thoroughly. Because writing requires respondents to retrieve a value from memory, while a list only requires them to recognize it, satisficing implies that even sincere respondents may not write down values they would select from a list.

Laurent et al. (1995) describe a brand's salience as its prominence in consumers' minds. For a brand of given salience, an easier question, such as recognizing the brand on a list instead of recalling it, yields higher measured awareness. Krosnick (1990) showed that attitudes important to respondents are more likely to be mentioned spontaneously and are more closely associated with political choices. Similarly, Fazio and Williams (1986) suggested that attitudes that are more readily accessible, measured by response speed, can also be more predictive of voting behavior. These two studies are closest to this paper: they relate the importance or accessibility of an attitude to choices, but neither compares how well the answers to open- and closed-ended questions align with them.

Advances in natural language processing have reduced costs for researchers to code for open-ended questions (Roberts et al. 2014, Singer and Couper 2017) and increased their use in economics (Haaland et al. 2025), including applications to climate change beliefs (Dechezleprêtre et al. 2025), perceptions of macroeconomic developments (Andre et al. 2022), economic narratives (Andre et al. 2026), the considerations behind investment decisions (Chinco et al. 2022), and policy preferences (Ferrario and Stantcheva 2022, Wekhof and Houde 2023). These studies use open-ended questions to measure what is top of mind without priming respondents through response options (Andre et al. 2026, Haaland et al. 2025), whereas structured lists of causal channels are easier to compare and require less effort (Andre et al. 2022); where both formats are used, they are validated against each other and found to be strongly correlated (Andre et al. 2026, 2022).

While the literature documents systematic differences in response patterns across question formats, the literature has only begun to study an economic criterion to select the appropriate question type: which elicitation format produces measures that align better with subsequent choices.[4] This paper applies this criterion to a fund choice and explains why the formats differ.

2.2 The value–action gap as a benchmark

A growing literature documents differences between stated values and subsequent actions, which environmental research calls the value–action gap (Blake 1999, Kollmuss and Agyeman 2002). Bertrand and Mullainathan (2001) showed that stated values are

[4]See, for example, Kraft (2024) and Filippini et al. (2024) for related evidence on the implications of open-ended measurement for knowledge gaps.

noisy but informative predictors of behavior. Choi and Robertson (2020) ask investors directly how well leading academic theories describe their financial beliefs and decisions. Chinco et al. (2022) validate closed-ended answers against choices: participants who tick that they considered the mean or volatility of stock returns respond more strongly to these parameters when allocating their portfolios, whereas those who tick the correlation with consumption growth do not respond to it. For sustainability, this gap is especially prominent in both consumption and investing (Anderson and Robinson 2022, Bethlendi et al. 2022, Nguyen et al. 2019, Vermeir and Verbeke 2006, White et al. 2019).

One explanation is that sustainable products tend to be more expensive: sustainable consumer goods carry price premiums (Nimon and Beghin 1999, Roheim et al. 2011), and sustainable funds charge higher fees (Raghunandan and Rajgopal 2022). Investors in socially responsible funds are willing to pay higher fees or accept lower returns (Bauer et al. 2021, Riedl and Smeets 2017).[5] For fees, the literature documents a gap of its own: limited attention to fees at the point of investment decision (Barber et al. 2005) and weak revealed fee sensitivity even when participants report caring about fees (Choi et al. 2010).

This literature measures values using closed-ended questions and focuses on frictions, norms, or information constraints, giving less attention to how the elicitation instrument itself shapes the measured prevalence of stated values and their alignment with choices. This paper uses the value–action gap to measure how well the answers to a survey question align with choices and shows that, for a given topic, the gap changes when the question switches from open to closed-ended, and that the size and direction of this change differ between topics.

2.3 Structural measurement and behavioral foundations

Behavioral economics provides micro-foundations for linking moral values elicited from surveys to a choice model. Akerlof and Kranton (2000) suggest that identity enters utility directly, so that actions that conform to the prescriptions of one's identity yield additional payoffs. Subsequent work shows that incorporating identity-based utility can explain choices across domains, including gender norms and household behavior (Bertrand et al. 2015) and political preferences (Grossman and Helpman 2021). In a broader social psychology context, the "theory of reasoned action" and its extension, the theory of planned behavior, hold that the intention to perform a behavior depends on attitudes toward that behavior (Ajzen 1991, Ajzen and Fishbein 1973). Unlike revealed preference approaches, where preferences are not stated before a choice but directly inferred from

[5]At the market level, Hartzmark and Sussman (2019) show that when Morningstar began publishing sustainability ratings for mutual funds, money flowed out of funds rated low and into funds rated high .

the choices themselves, the theory of reasoned action elicits attitudes first and uses them to predict behavior.

Moving from an open to a closed-ended question shows respondents the values to choose from, which turns the task from recall (open) into recognition (closed). In the survey, a value that is not written may simply not have come to mind; a listed value is in front of the respondent, so not ticking it is a clearer "no." Models of limited attention make the same point about choices: an option that is not chosen tells us something about preferences only if the decision maker paid attention to it (Manzini and Mariotti 2014, Masatlioglu et al. 2012).

Weak links between stated values and behavior can arise at several stages. When the choice is made, cognitive noise can influence the mapping from valuation to decision (Enke and Graeber 2023); when values are measured, elicitation noise can weaken the measured relationship between values and choice (Gillen et al. 2019). Writing an answer also takes more effort than ticking a box, which Haaland et al. (2025) discuss as a source of measurement error and nonresponse. In this paper, measurement matters through selection rather than noise: the question determines which convictions become visible before the fund choice is made. Open-ended answers come from respondents with stronger convictions but miss respondents who tick the value on a list and still act on it more often than those who do neither. Gillen et al. (2019) also elicit the same preference more than once, but to correct for noise: they treat the answers as noisy copies of one measure. Here, the two formats are not noisy copies; they select different parts of the conviction distribution.

To separate this selection from the choice itself, the paper models the survey answers and the fund choice together. Handel and Kolstad (2015) combine survey answers and choices in one model to study frictions in the choice: they incorporate survey measures of consumer information and perceived hassle costs into an insurance choice model to separate the influence of these frictions on plan choices from that of risk aversion. Hybrid choice models do the same for attitudes: they treat an attitude as a latent variable that the survey answers measure with noise and that enters the utility of the alternatives (Ben-Akiva et al. 2002). This paper's model belongs to this class. Its answers are thresholded readings of the latent value, as in hybrid choice models with ordered indicators (Daly et al. 2012), and the latent value enters the utility of the funds directly: the stronger the conviction, the more its value counts in the choice, as in social psychology, where values guide choices more when they are activated and central to the self (Verplanken and Holland 2002). Closest to this paper, studies of stated attribute non-attendance ask respondents whether they ignored an attribute and let the answers inform whether, or how much, it counts in the choice (Hess and Hensher 2013, Hole et al. 2013). There, respondents answer after a series of choices, so the answers may reflect the choices; here,

the value questions come before a single choice. Within these models, what is new is that the randomized question format sets the threshold, so that the two formats reveal different parts of the same latent value distribution.

3 Survey and experimental design

The survey was conducted between October and November 2021 with 3,059 participants in the German-speaking region of Switzerland.[6] A professional survey company recruited participants from a large panel of households and incentivized them to participate in the survey with a payment.

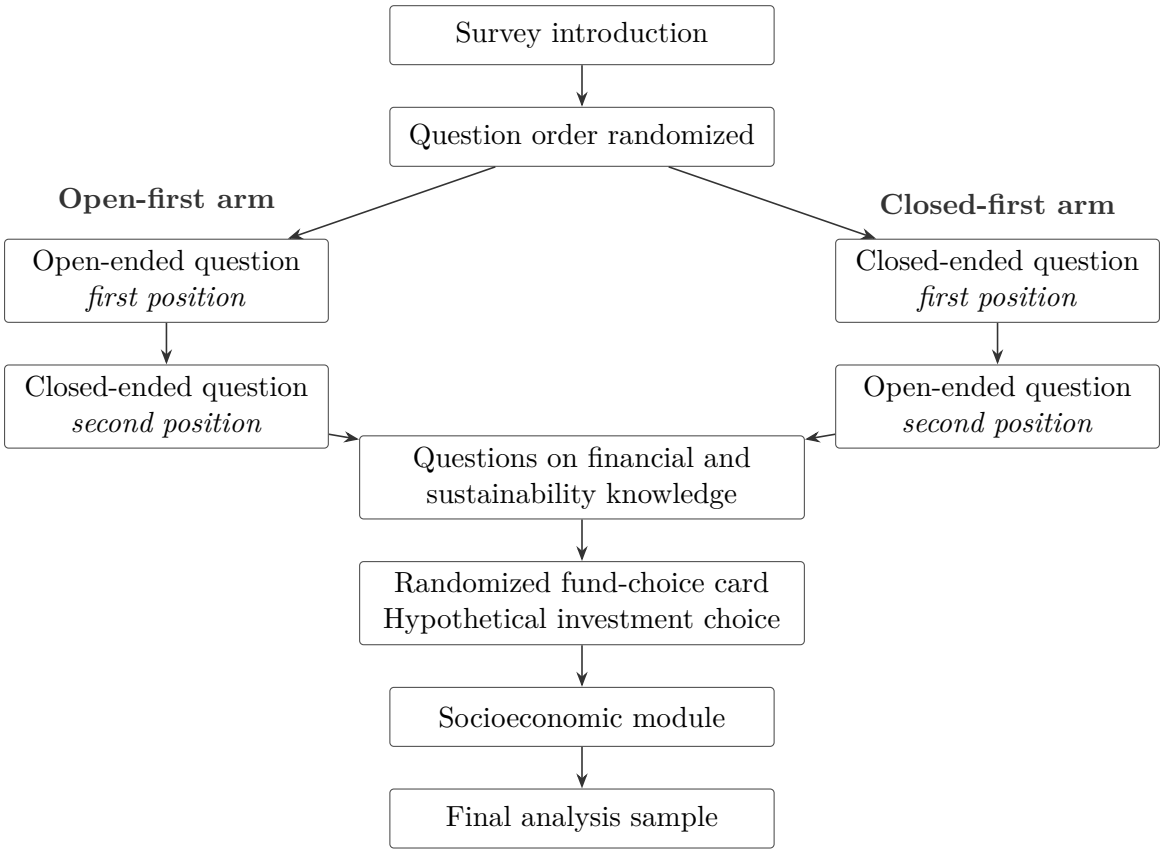
The survey company screened participants to include only respondents with basic investment experience. Participants could participate in the survey if they had invested in the Swiss voluntary pension plan (in 2019, about 62% of the Swiss working population had participated).[7] I conducted a pretest in October 2021 with 360 households and, based on the results, simplified the questionnaire and refined the closed-ended options using themes that emerged from the open-ended answers, following Schuman and Presser (1979), Schuman and Scott (1987).

The survey experiment consisted of two main parts, which will be described in more detail in the next subsections. Figure 1 shows the study design: after starting the survey, the survey elicited investment values. No part of the survey (including the invitation) prior to the value elicitation, including the survey invitation, mentioned sustainability. Half of the sample answered an open-ended question followed by a closed-ended equivalent question; both questions asked about the attributes of an investment product that were important to the respondent. The other half of the sample first answered a closed-ended question and the open-ended question second; the order of the closed-ended options was randomized. After the value elicitation and a block of questions on financial and sustainability knowledge, which was the same in both arms, respondents participated in a hypothetical investment experiment, followed by a socioeconomic survey. Because the order is randomized, the first answers of the two groups compare the two formats cleanly: each group answers its first question without having seen the other format, whereas the second answer may be influenced by the first question.

[6]The study is based on the same survey used by Filippini et al. (2024) and focuses on a different section of the survey. The part of the survey used in this paper was registered after data collection in the AEA RCT Registry AEARCTR-0017723.

[7]For more information, see https://www.bfs.admin.ch/bfs/de/home.html. Official statistics are only available for the Swiss working population, while the sample also includes pensioners, making a comparison to the general population challenging. According to data on Swiss residents actively contributing to the Swiss voluntary pension plan, the average age within this group is 43 years, 43% are women, and 50% hold a university degree. Excluding pensioners from the sample, the mean age was 45 years, 46% women and 55% held a university degree. In the full analysis sample, the mean age is 49 years, 45% are women, and 53% hold a university degree (Appendix Table D.1).

Figure 1: Study design



Note: Of 4,233 entrants, 3,059 remained in the final sample: 1,565 open-first and 1,494 closed-first. Appendix Table D.2 reports break-off at each stage.

Table 1 summarizes the content of the two response formats; Appendix D.4 reproduces the complete wording. The closed-ended question offers both fees and sustainability as options, and the order of all options was randomized, so a non-selection of these topics follows an explicit exposure regardless of their position in the list.

Table 1: The elicitation contrast (translated excerpts)

Open-ended (unaided recall)	*Closed-ended (aided recognition)*
"What criteria would be important to you when choosing a fund? Please write a short text with about three sentences." *No investment topics were named.*	"What criteria would be important to you when choosing a fund? (multiple options possible)." ☐ Expected return ☐ **Management fees** ☐ **Sustainability of the included companies** *seven further topics plus "other"*

Note: Both formats use the same CHF 10,000 investment scenario. Appendix D.4 reproduces every order variant and the complete closed-ended question verbatim. Bolding of the two analysis topics is added here and was not shown to respondents. The options appeared in randomized order; the excerpt shows one ordering.

The majority of open-ended responses in the survey were well-written and thoughtful; on average, an answer has 13 words in 1.6 sentences (Appendix D.2). For example, one respondent wrote: *"Environmentally friendly investment, good profit, transparent."* This sentence contains the keyword "environmentally," classifying the response into the topic "sustainability." Additionally, the answer contains the words "profit," which adds the "return" topic, and "transparent," which adds the topic "transparency." I allocated topics to the open-ended responses using a semi-automated method developed by Wekhof and Houde (2023), which was also employed in Filippini et al. (2024). A 500-response hand-coded validation sample and a large-language-model coding both yield a Cohen's $\kappa \geq 0.95$ for fees and sustainability (Appendix D.3).

Later in the survey, respondents chose between two hypothetical mutual funds that were identical in past performance and risk but differed in sustainability and annual management fees (1% vs. 1.5%). Both funds excluded the weapons and coal mining sectors and scored well on at least one ESG dimension, but only the sustainable fund had an explicit sustainable investment objective (i.e., to include companies "contributing to a sustainable future"). For short, I call the other fund conventional; it is less sustainable, not unsustainable. Overall, 67% of respondents chose the sustainable fund.[8]

In the choice stage, respondents were randomly assigned to one of four choice cards. The control card shows no sustainability rating. The three treatment cards add a sustain-

[8]A sustainable choice could also reflect expected returns (Hartzmark and Sussman 2019). Before the choice, respondents stated how they expected the returns of sustainable equity funds to compare with those of conventional equity funds: 42% expected lower returns, 32% identical returns, 17% higher returns, and 9% did not know; the distribution does not differ significantly between the two question-order arms ($p = 0.80$). Respondents who expected higher returns chose the sustainable fund no more often than those who expected lower returns (65% versus 69%).

ability rating to both funds, always A for the sustainable fund and C for the conventional fund; they differ in who issues the rating (the government or the bank) and in what it covers (all three ESG areas or climate mitigation only). The cards were assigned after value elicitation, and their variation in subsequent choice enters the structural estimation. Appendix Figures D.1 and D.2 reproduce all four cards (translated from German).

3.1 Break-off and hypothetical choice

Surveys with open-ended questions typically face two problems: break-offs, when respondents quit the survey after encountering an open-ended question, and poor answer quality, often referred to as "non-response" (Hadler 2025). Break-off at the open-ended question is 10.4% in first position and 13.8% in second position, compared with 3.9% and 1.9% at the corresponding closed-ended questions. Women, respondents without a university degree, and less wealthy respondents break off more often, but the differences are small (Appendix Table D.3, which uses characteristics recorded in a 2024 update of the panel). Among respondents who answered the open-ended question first, 6.7% gave non-responses such as "do not know" or "no interest," fewer than in most earlier studies (Appendix D.2). Because every respondent encounters both formats, final completion is only slightly higher in the open-first arm (74.4% versus 70.2%).

Some studies found hypothetical bias affecting the magnitude of willingness to pay in stated choice experiments (Fifer et al. 2014, Haghani et al. 2021). However, Hainmueller et al. (2015) suggest that the structural determinants of preferences align with real-world settings. Several studies have shown that hypothetical survey responses correlate with real-world behaviors (Fehr et al. 2024, Funk 2016, Hainmueller et al. 2015). Similarly, Engler et al. (2025) found that willingness to pay for sustainable investments was not significantly higher in non-incentivized than in incentivized experiments. This means that the value-action gap from this study could be larger in a real-world setting, but the underlying factors are likely similar.

3.2 Other surveys

To validate the patterns found for fees and sustainability with other values and outcomes, I use three surveys that ask the same respondents about two competing values, first in an open question and then in a closed-ended question, and record an outcome that both values can influence: the third wave of Andre et al. (2022), in which US households predicted inflation after a hypothetical shock and reported the propagation channels they had in mind; the Social Identities and Cleavage Formation Survey (Bornschier et al. 2024), here called the Zurich survey, in which voters in four European countries described the people who are like them, rated how close they feel to their country's people and to cosmopolitans, and named the party they would vote for; and a Swiss survey from 2018

that asked similar questions (Bornschier et al. 2021). I use a survey only if its closed-ended question records the respondent's own concern or identification, not an assessment of the alternatives. Table 2 summarizes them. All three are publicly available, from the survey organizations or in the authors' replication packages. Section 7.3 reports the results, and Appendices E.3 and E.4 give the details.

Table 2: Surveys with open and closed questions on the same considerations

Survey	Respondents	N	Value	Open/closed	Outcome
This survey	Swiss retail investors	3,059	Fees	16.5 / 53.7	Fund choice
randomized	2021		Sustainability	35.4 / 41.0	
Andre et al. (2022)	US households	1,607	Cost channel	13.1 / 57.7	Forecast
wave 3, paired	three scenarios		Demand channel	15.9 / 56.2	
Zurich survey	Voters, four countries	7,886	National identity	2.3 / 72.4	Vote
paired	2020–21		Cosmopolitan identity	6.8 / 40.7	
Swiss survey	Swiss voters	656	National identity	7.8 / 86.1	Vote
paired	2018		Cosmopolitan identity	15.1 / 63.3	

Note: The less salient value, to which the list adds more respondents per writer, comes first. Open/closed: percentage of respondents who report the value in the open-ended and in the closed-ended question, for this survey in first position. Closed-ended reports are a closeness of at least 7 on a 1–10 scale (Zurich survey) and a rather or very close identification (Swiss survey). Randomized: different, randomly assigned respondents answer the two formats; paired: the same respondents answer both, the open question first. Outcomes: the hypothetical fund choice; the inflation forecast (rising follows the cost channel, falling the demand channel); a vote for the left bloc (identity surveys).

4 Descriptive results: Open vs. Closed-ended questions

This section shows the topic shares for the open-ended and closed-ended responses. While both formats consistently identified *risk* and *return* as the most frequently reported topics, closed-ended responses yielded more topics per person and higher frequencies for most listed items. However, open-ended answers revealed a broader range of themes, including ones not captured by the pre-defined closed-ended items.

The topic rankings were similar at the top and bottom but diverged in the middle. In first-position responses, sustainability ranks third under open-ended elicitation but sixth on the closed-ended question, behind fees, transparency, and trust in the bank. Of the 45 pairs formed by the ten topics asked in both formats, eight swap order between the formats in the estimates, and only the three involving sustainability remain significant after a Bonferroni correction.[9] The most relevant of these reversals is between sustainability and fees, the two attributes of the fund choice: sustainability leads when respondents write, fees lead when they tick. The reversals show that the list does not raise every topic by the same amount: it adds far more respondents to fees, transparency, and trust in the

[9]With 45 pairs, some reversals could appear by chance. The Bonferroni correction guards against this by testing each pair at the 5% level divided by 45.

bank than to sustainability, and a common shift of all topics is rejected (Appendix E.1; Figure E.1 shows the full rankings).

Table 3 compares how often the two topics relevant to the choice experiment, sustainability and management fees, surface in respondents' answers under the four combinations of question format (open- vs. closed-ended) and ordering (asked first vs. second). Comparing first answers, 41.0 percent select sustainability in the closed-ended question and 35.4 percent write about it, whereas 53.7 percent tick fees and only 16.5 percent write about them. This indicates that the sustainability topic is relatively salient among respondents, as suggested by Gibson Brandon et al. (2023), and not a niche subject. Once respondents in the open-first group proceeded to the closed-ended question, the prevalence of the topics rose sharply to 51 percent for sustainability and 53 percent for fees. Seeing the closed-ended question first raises later open mentions of seven of the ten topics by 2 to 12 percentage points (Appendix Table E.1). Fee mentions rise from 16.5 to 25.8 percent, whereas sustainability mentions remain similar, at 35.4 and 33.7 percent, as do those of risk and ETFs.

A closed-ended endorsement takes one of two observed forms. An endorsement is *equivalent* if the respondent selects the topic in the closed-ended question and also mentions it in the open-ended answer. It is *non-equivalent* if the respondent selects the topic in the closed-ended question but does not mention it openly.

Almost two-thirds of closed-ended sustainability answers are supported by a written explanation (64–66 percent), whereas for fees, the majority are not (only 28–41 percent are equivalent). Many investors check the "management fees" box when it is offered, but do not spontaneously write about cost awareness. In contrast, respondents who claim to care about sustainability are equally likely to elaborate on it in an open-ended question, with or without priming by a preceding closed-ended question.

Table 3: Topic shares, endorsement equivalence, and value–action gaps

	Sustainability		Fees	
Question order	Open first	Closed first	Open first	Closed first
Open-ended (%)	35.4	33.7	16.5	25.8
Closed (%)	51.4	41.0	53.4	53.7
Equivalent (% of closed-ended question)	63.6	66.4	28.2	41.3
Non-equivalent (% of closed-ended question)	36.4	33.6	71.8	58.7
Value–action gap (%)				
Open-ended reporters	17.0	16.7	48.1	58.2
Closed-ended reporters	19.0	16.0	59.2	62.6
Respondents	1,565	1,494	1,565	1,494

Note: This table shows, by randomized question order, the share of respondents who report each topic in each format; the split of closed-ended endorsements into those also mentioned in the open-ended answer (equivalent) and those not mentioned (non-equivalent); and the value–action gap, the share of reporters who choose the fund that goes against the topic. First answers are the open-ended rows in the open-first columns and the closed-ended rows in the closed-first columns; their value-action gap differences are 14.5 points for fees (SE 3.5) and −1.0 for sustainability (SE 2.2).

The last two rows of Table 3 relate these reports to the fund choice, and report the value–action gap: the share of respondents who report a topic but choose against it, that is, the high-fee fund for fee reporters and the conventional fund for sustainability reporters. Its level depends on the choice offered, a fee premium of 0.5 percentage points for more sustainability, which is the same in all comparisons. Considering one value at a time, sustainability answers align with choices about equally in both formats: 17.0% of open-ended sustainability mentions and 16.0% of closed-ended sustainability endorsers choose the conventional fund. Fees look different. Under closed-ended elicitation, 62.6% of investors who endorse low-fee importance still choose the high-fee fund; under open-ended elicitation, 48.1% do. The second answers show the same pattern: the sustainability gap stays between 16 and 19 percent, whereas fee writers who saw the list first have a gap of 58.2 percent, similar to the closed-ended gaps.

These reporting differences establish how open vs closed-ended questions influence which values respondents express. The next section examines how those reported values relate to subsequent fund choice in OLS regressions and adds the competing value as a control variable.

5 Stated Values and Fund Choice

This section relates the reported values to the hypothetical fund choice with OLS regressions. OLS allows estimating the association of the target value with choice while controlling for the competing value mentions. Section 5.1 reports the estimates and the value–action gaps they imply when the competing value is not reported; Sections 5.2

and 5.3 examine question order and equivalent answers, and Section 5.4 reports robustness checks.

5.1 OLS estimates

Let $m \in \{O, C\}$ denote question format (open or closed-ended), $\omega \in \{OC, CO\}$ question order (open-then-closed or closed-then-open), and $j \in \{S, F\}$ the topic. Because every respondent answers both questions sequentially, each question order gives one open and one closed-ended answer per respondent. This yields four combinations (m, ω) of format and order. In the order OC, respondents first give the unprimed open answer (O, OC) and then the closed-ended answer (C, OC); in the order CO, they first give the closed-ended answer (C, CO) and then the primed open answer (O, CO). For each combination, I estimate an OLS regression:

$$y_i = \alpha^{m,\omega} + \beta_S^{m,\omega} T_i^{m,\omega,S} + \beta_F^{m,\omega} T_i^{m,\omega,F} + \varepsilon_i. \tag{1}$$

where y_i is a binary indicator that takes the value of one when choosing the sustainable high-fee fund, and $T_i^{m,\omega,j}$ indicates whether respondent i reports topic j in the survey. Both topics enter the regression separately, so each coefficient is an adjusted difference in choice probability, net of the other topic's report. These differences are measured in percentage points, because the linear probability model acts as an additive summary of the binary outcomes.

Columns (1) and (3) of Table 4 report how mentioning each topic in the open- or closed-ended response is associated with choosing the sustainable fund. The sustainability topic is positively and statistically significantly associated with sustainable fund selection in both question formats, with nearly identical coefficients (0.26 open-ended, 0.28 closed-ended).

For fees, the association is negative and stronger in magnitude with the open-ended format: respondents mentioning fees in the open-ended response are 18 percentage points less likely to choose the high-fee sustainable fund, compared to a 12-percentage-point lower likelihood with closed-ended elicitation. The regression estimates therefore suggest a stronger association for open-ended fee responses, although the 5.7-percentage-point difference between the coefficients is not statistically significant ($p = 0.15$).

Because the fund choice trades sustainability against fees, a respondent who reports both values must trade them off, so a choice against the target value is not required to reveal an inconsistency. I therefore measure each value's gap net of the competing value, by controlling for this variable in the regression. In eq. (1), with the target value reported

Table 4: OLS associations between stated values and fund choice

Specification	(1)	(2)	(3)	(4)	(5)	(6)
	Two topic indicators				Equivalence indicators	
Measure	Open first	Open after closed	Closed first	Closed after open	Closed first	Closed after open
Panel A. Two-indicator coefficients						
Fees	−0.177***	−0.121***	−0.120***	−0.134***		
	(0.033)	(0.028)	(0.023)	(0.022)		
Sustainability	0.257***	0.230***	0.278***	0.316***		
	(0.022)	(0.023)	(0.022)	(0.023)		
Panel B. Equivalent and non-equivalent endorsements						
Fees (equiv.)					−0.169***	−0.234***
					(0.031)	(0.034)
Fees (non-equiv.)					−0.095***	−0.089***
					(0.027)	(0.024)
Fees (write only)					−0.106*	0.014
					(0.062)	(0.092)
Sustainability (equiv.)					0.316***	0.349***
					(0.024)	(0.025)
Sustainability (non-equiv.)					0.239***	0.253***
					(0.033)	(0.031)
Sustainability (write only)					0.130***	0.096
					(0.050)	(0.075)
Intercept	0.594***	0.631***	0.628***	0.565***	0.622***	0.561***
	(0.017)	(0.017)	(0.020)	(0.022)	(0.022)	(0.022)
Observations	1565	1494	1494	1565	1494	1565
R^2	0.093	0.069	0.100	0.131	0.112	0.148

Note: This table shows OLS estimates of eq. (1); the outcome is choosing the sustainable high-fee fund. Columns (5)–(6) split closed-ended endorsements into equivalent (also mentioned openly) and non-equivalent ones. Write only: written but not ticked (97 and 43 sustainability writers, 54 and 22 fee writers in columns 5 and 6); respondents who neither write nor tick a topic are the reference group. No covariates or fixed effects; HC1 standard errors in parentheses. With both reports, R^2 is about the same for written and ticked first answers (columns 1 and 3), as the per-topic differences in Section 5.4 offset. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

and the competing topic value indicator set to zero, the predicted value–action gaps are

$$G_F^{m,\omega} = \widehat{\alpha}^{m,\omega} + \widehat{\beta}_F^{m,\omega}, \qquad G_S^{m,\omega} = 1 - \widehat{\alpha}^{m,\omega} - \widehat{\beta}_S^{m,\omega}. \tag{2}$$

For fees, the gap is choosing the high-fee (sustainable) fund, which is the fitted probability of $y_i = 1$; for sustainability, it consists of the non-sustainable (low fee) option, which is one minus that probability. The competing topic's coefficient drops out because its indicator is zero.

These OLS-implied gaps give a different comparison than the coefficients. The OLS estimates imply a fee gap of 41.7% under open-ended elicitation and 50.8% under closed-ended elicitation. This 9.1-percentage-point increase is statistically significant at the 5% level. For sustainability, the gap falls from 14.9% to 9.4%, a 5.6-percentage-point decrease that is also significant at the 5% level. Both formats are evaluated in first position of the question randomization.

These gaps change more than the coefficients because each gap also includes the intercept, the OLS-implied sustainable-fund share among respondents who report neither topic. This baseline can differ between formats (0.594 and 0.628) because the group reporting neither topic is itself defined by the format: the closed-ended baseline group excludes respondents who would recognize fees or sustainability on the list. The change in the fee gap is therefore the change in the intercept plus the change in the fee coefficient: $9.1 = 3.4 + 5.7$ points.

The raw gaps in the same groups move in the same way: the fee gap rises from 39.7% to 50.5%, while the sustainability gap falls from 15.7% to 9.9% (Appendix Table C.1). Adding the competing attribute value thus changes the trade-off between the two question types, but differently for each of the two topics: the closed-ended question adds weakly held fee endorsements, which raises the fee gap, and not ticking fees is a stricter screen than not writing about them, which lowers the sustainability gap. Among sustainability reporters, the share who also report fees rises from 12 percent with the open question to 55 percent with the list, and these respondents choose the low-fee fund more often. Netting out the fee report in OLS therefore puts the closed-ended sustainability gap 6.6 points below its one-value level in Table 3, against 2.0 points for the open-ended gap.

5.2 Question order

Columns (2) and (4) use responses given second: i.e., open-ended answers after the closed question has already been answered and closed answers after the open question; these comparisons may combine different effects like priming, deliberation, and carryover.

Question order does not affect the sustainability coefficient compared to column (1); but the fee coefficient in the open-ended response falls in magnitude to -0.12, approximately matching the closed-ended estimate, while priming increases the frequency of open-ended fee responses (from 17% to 26%). In contrast, closed-ended answers after writing have slightly larger coefficient magnitudes for both sustainability and fees (column 4). Overall, priming increases the frequency of less salient topics, like fees; the estimates also suggest a weaker choice association, but that coefficient change is not statistically significant. For more salient topics like sustainability, neither question-order comparison has a statistically detectable impact on the coefficient.

5.3 Equivalent and non-equivalent answers

I split the closed-ended indicator into the two observed forms defined in Section 4, equivalent and non-equivalent endorsements, and add a write-only control:

$$\begin{aligned} T_i^{\text{eq},C,\omega,j} &= \mathbf{1}\{T_i^{C,\omega,j} = 1,\ T_i^{O,\omega,j} = 1\}, \\ T_i^{\text{neq},C,\omega,j} &= \mathbf{1}\{T_i^{C,\omega,j} = 1,\ T_i^{O,\omega,j} = 0\}, \\ W_i^{\omega,j} &= \mathbf{1}\{T_i^{C,\omega,j} = 0,\ T_i^{O,\omega,j} = 1\}. \end{aligned}$$

The regression is:

$$\begin{aligned} y_i = \alpha^{\omega} + \sum_{j \in \{S,F\}} \Big[& \beta_j^{\text{eq},\omega} T_i^{\text{eq},C,\omega,j} \\ & + \beta_j^{\text{neq},\omega} T_i^{\text{neq},C,\omega,j} + \beta_j^{W,\omega} W_i^{\omega,j} \Big] + \varepsilon_i. \end{aligned} \tag{3}$$

Columns (5) and (6) estimate equation (3) using closed-ended responses but distinguish between equivalent and non-equivalent endorsements, i.e., whether the topic also appeared in the respondent's open-ended response. When the respective question comes first, 91 and 92 percent of open fee and sustainability mentions are also selected on the closed-ended question, but only 41 and 66 percent of closed-ended fee and sustainability selections are also mentioned openly (Appendix Table E.3).

In both columns, the estimates show a stronger association with choice for equivalent than for non-equivalent endorsement. For sustainability, the equivalent coefficient is 0.316 (column 5), compared to 0.239 for non-equivalent; for fees, the equivalent coefficient is -0.169, compared to -0.095. Non-equivalent endorsements, which appear only on the list, are thus weaker signals of the choice than equivalent ones, but they still carry information: in columns (5) and (6), their coefficients are 38 to 76 percent as large as the equivalent ones, and all four are significant at the 1% level. Respondents who write a topic without ticking it act on it less than those who do both: the write-only coefficients are smaller in magnitude than the equivalent ones, and only one of the four is significant

at the 5% level. In the estimates, priming (in column 5, the open answer that defines equivalence is given after the closed-ended question) mainly affects the fees topic by decreasing the magnitude of the equivalent coefficient and thus reducing the difference between equivalent and non-equivalent endorsements. For sustainability, the coefficients in columns (5) and (6) are similar.

5.4 Robustness

I conducted several robustness checks of the gap changes in Section 5.1. Both changes are significant at the 5% level, whether computed from the OLS fits or from the raw gaps (Appendix Table C.1). They become larger without the 17% of respondents who expected sustainable funds to earn higher returns: the raw fee gap then rises from 35.4% to 49.6%, and the raw sustainability gap falls from 16.1% to 8.9%. Coding the open-ended answers with a large language model instead of the keyword dictionary changes the open-ended gaps by less than one percentage point (Appendix D.3). How reports relate to choice does not differ significantly between the two question orders or among the three choice cards with a sustainability rating (Appendix B.6).

The gap comparisons use respondents from different arms, which differ in completion (74.4% versus 70.2%; Section 3.1). If the excess open-first completers are removed in the least favorable way, assuming they are spread proportionally across reporting groups, the raw fee increase of 10.8 points remains at least 8.5 points, so differential completion cannot explain it. The raw sustainability decrease of 5.9 points remains significant at the 5% level unless the excess completers who wrote about sustainability chose the conventional fund at more than twice the rate of other sustainability writers, a rate that no observed group of writers comes close to (Appendix D.1).

For each topic, the format with the smaller gap also explains more of the variation in fund choice (R^2): on their own, fee reports explain 2.7 percent of it when written and 1.4 percent when ticked, and sustainability reports 7.4 and 8.4 percent. Unlike the gap, the explained share also depends on how many respondents a format counts as reporters.

Further checks, reported in Appendix F, relate the reports to respondents' socioeconomic characteristics. Regressions of topic mentions on these characteristics reveal consistent patterns across question types: women, higher education, and altruism are positively associated with sustainability, while age and financial literacy are important predictors for fees. An additional specification uses the sustainable fund choice as the outcome variable and the background variables as independent regressors; its coefficients align with those of the topic regressions, most closely with the open-ended ones. Finally, a multinomial logit regression comparing equivalent and non-equivalent closed-ended responses with the baseline of mentioning no topic reveals that equivalent answers are more

strongly associated with background traits, supporting the interpretation that equivalent answers reflect more deeply held values.

The reduced-form comparisons measure the total change in the value–action gap from open to closed-ended questions, but not its two parts. The first part is target broadening: the closed-ended question adds respondents with weaker convictions to the group that reports the target value, which raises the gap. The second is competing-value screening: the closed-ended question removes respondents who would recognize the competing value from the group that does not report it, which lowers the gap. Separating the two parts would require a gap in which only one of the two values changes format, for example a closed-ended target value with an open-ended competing value, both given as first answers. No respondent can give both as first answers: in the data, such a combination always includes a second answer, which may be influenced by the preceding question. The structural model in Section 6 describes both first-position reporting rules for the same population and can therefore separate the two parts.

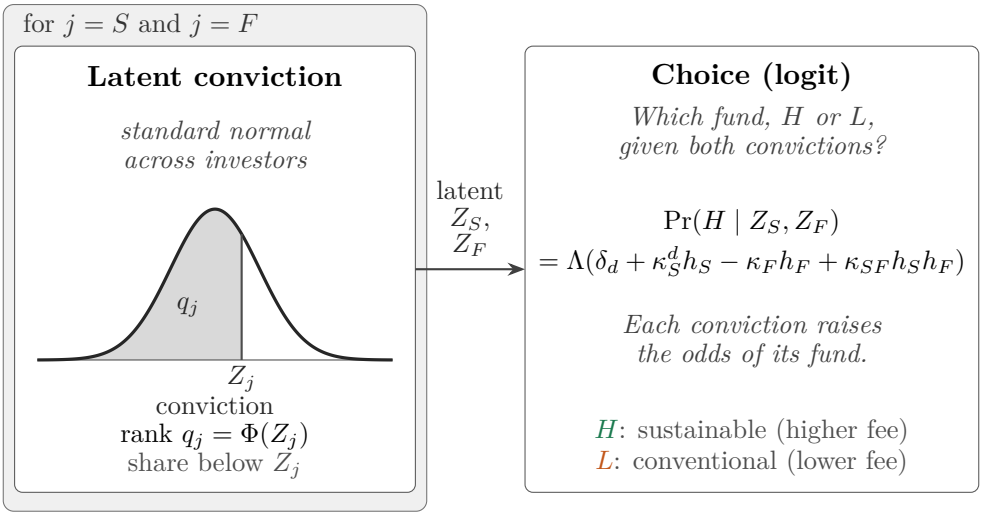
6 Theory: Values-Based Utility

This section develops a model linking the strength of investors' values to the values reported in the survey and to the fund they choose. It is a random-utility model in which a stronger conviction for a value raises the utility of the fund that the value favors. Each investor has a latent conviction for sustainability and for fees. The survey reveals a conviction only above a reporting threshold, which is lower for selecting a closed-ended answer than for writing.[10] Each component follows existing models, as the subsections below note; what is new is the measurement, in which the same value question, asked in two randomized formats, places two thresholds on one latent conviction. The model thereby (i) explains the reduced-form results, because open and closed answers capture different upper segments of each topic's conviction ranking, and (ii) shows how the two margins in Table 5 depend on the question format and on the salience of each attribute.

The model has a latent conviction for each value and a choice step (Figure 2). The choice is a logit regression of the fund choice on the two convictions and their product: each value's conviction raises the probability of choosing the fund that the value favors, and the value's weight measures by how much. The competing value thus acts as a second regressor, as in the regressions of Section 5. Section 6.3 adds the reporting thresholds.

[10] In the terminology of hybrid choice models (Ben-Akiva et al. 2002, Walker 2001), conviction is the latent variable, the open- and closed-ended answers are its indicators, and the fund choice is a logit model in which the latent variables enter the utility of the alternatives. Section 2.3 describes what the model shares with them.

Figure 2: Latent conviction and the choice step of the model



Note: For each value j, conviction Z_j is standard normal across investors, and the rank $q_j = \Phi(Z_j)$ is the share of investors with weaker conviction. The choice step is the logit of eq. (6), in which each conviction enters through its index $h_j = h_j(Z_j)$ of eq. (4). The convictions are not observed; the survey answers reveal them only above the reporting thresholds of Section 6.3. Drawn for $Z_j = 0.6$; $\Lambda(z) = 1/(1 + e^{-z})$.

6.1 Latent conviction

Individual i chooses between the sustainable high-fee fund H and the conventional low-fee fund L; let $Y_i = 1$ denote choosing H. For each topic $j \in \{S, F\}$, the individual has a *latent conviction* $Z_{j,i}$, which governs how strongly the moral/identity payoff from attribute alignment enters the decision. The underlying idea is that valuing an attribute increases the utility from making a value-consistent choice, conceptually related to Akerlof and Kranton 2000. A value-consistent choice is $C_{S,i} = Y_i$ for sustainability, and $C_{F,i} = 1 - Y_i$ for fees.

As in hybrid choice models (Walker 2001), conviction is standard normal across investors, $Z_{j,i} \sim N(0, 1)$. Its location and scale are normalizations: with ordinal answers and a single choice, only the ranking of convictions and the shape of the choice along this ranking are identified. The model estimates that shape with one skew parameter g_j per value, which bends the scale on which conviction enters the choice, Tukey's g-transformation (Hoaglin 1985):

$$h_j(Z) = \frac{e^{g_j Z} - 1}{g_j}. \tag{4}$$

With $g_j = 0$, the index is the conviction itself. With $g_j < 0$, it is bounded above at $-1/g_j$: beyond some point, a stronger conviction adds little. With $g_j > 0$, it stretches

the top of the ranking. Appendix B.4 compares other scales. Let

$$q_{j,i} = \Phi(Z_{j,i}) \in [0, 1]$$

denote respondent i's quantile rank from weakest to strongest conviction, where Φ is the standard normal distribution function. As a quantile rank, $q_{j,i}$ is uniformly distributed on $[0, 1]$ within each topic, so the top share r of respondents has ranks in $[1 - r, 1]$.

6.2 From conviction to fund choice

The fund choice involves both values: the sustainable fund charges the higher fee, so the two values pull in opposite directions. Respondent i chooses between the sustainable high-fee fund H and the conventional low-fee fund L on a randomly assigned choice card. I pool the three ESG-rating cards into one group, because the reports' coefficients in regressions of fund choice do not differ significantly across these cards (Appendix B.6.2). The card indicator therefore distinguishes the control card ($D_i = 0$) from any of the three ESG-rating cards ($D_i = 1$). Conditional on the two convictions and the randomized choice card $D_i = d$, the utility difference is

$$\begin{aligned} U_i(H) - U_i(L) &= V_d(q_{S,i}, q_{F,i}) + \epsilon_i, \\ V_d(q_S, q_F) &= \delta_d + \kappa_S^d h_S(q_S) - \kappa_F h_F(q_F) + \kappa_{SF} h_S(q_S) h_F(q_F), \end{aligned} \tag{5}$$

where $h_j(q)$ is short for the index at rank q, $h_j(\Phi^{-1}(q))$, and ϵ_i is logistic, the difference of two independent type-I extreme-value shocks, as in the binary logit (McFadden 1974). The probability of choosing the sustainable fund on card d is therefore

$$\Pr(Y_i = 1 \mid q_{S,i}, q_{F,i}, D_i = d) = \Lambda(V_d(q_{S,i}, q_{F,i})), \tag{6}$$

where $\Lambda(z) = 1/(1 + e^{-z})$. The intercept δ_d is the choice of an investor with median convictions: a conviction above the median raises the utility of the fund its value favors, and one below lowers it. The weights κ_S^d and κ_F measure how much one standard deviation of conviction moves the log odds near the median, where the index has slope one. The product term lets each value's weight depend on the other value's conviction. The sustainability weight and the intercept depend on the card, because the rating cards show the funds' sustainability; the fee weight does not, because every card shows the same fee difference.[11]

The choice is the classic revealed-preference setting of a regression of the choice on the attributes: the investor trades off the funds as the card shows them, a sustainable

[11] Estimated separately by card, the fee weight does not differ: the rating-card weight is 0.84 times the control-card weight, and the criterion improves by 0.7 (Appendix B.4).

investment objective and, on the rating cards, a better rating against a lower fee, and each attribute counts more the stronger the investor's conviction for the corresponding value. Given the convictions, this is the conditional logit of McFadden (1974), in which the convictions weight the funds' attributes; averaged over the convictions, it is a mixed logit (McFadden and Train 2000). With more than two options, a conviction raises the utility of the option its value favors, as a voter's issue position raises the utility of the closest party in the spatial model of voting (Alvarez and Nagler 1998). The value questions come before the card, so the answers measure conviction, not the trade-off itself. With the same fee difference on every card, the model estimates the weights once for each card design; how investors with the same convictions differ is the logistic shock ϵ_i.

Averaging eq. (6) over the competing value's conviction and over the randomly assigned cards, denoted E_D, gives the probability of a value-consistent choice at each conviction rank of the target value.[12] At sustainability rank q and fee rank u, the index of eq. (5) is

$$V_D(q,u) = \delta_D + \kappa_S^D h_S(q) - \kappa_F h_F(u) + \kappa_{SF} h_S(q) h_F(u),$$

and the two curves are

$$\begin{aligned} P_S^{VC}(q) &= \mathrm{E}_D\left[\int_0^1 \Lambda(V_D(q,u))\,du\right], \\ P_F^{VC}(q) &= 1 - \mathrm{E}_D\left[\int_0^1 \Lambda(V_D(u,q))\,du\right]. \end{aligned} \tag{7}$$

Here $P_S^{VC}(q) = \Pr(Y = 1 \mid q_S = q)$ and $P_F^{VC}(q) = \Pr(Y = 0 \mid q_F = q)$. Both curves rise with the target value's rank. Their shape depends on the weights, on the skews, on the competing value, and on the card mix: a negative skew bounds the index above, so the curve flattens among the strongest convictions.

6.3 Reporting thresholds

Let r_{jm} denote the share of respondents who report value j in question format m, with $m = C$ for the closed-ended list and $m = O$ for the open-ended question. The model assumes that reporting follows conviction, as indicators follow a latent variable in hybrid choice models with ordered answers (Daly et al. 2012); the new element is that the question format sets the threshold. Ranking respondents from the weakest to the strongest conviction, $q_j \in [0,1]$ as in Section 6.1, the closed-ended endorsers are the share r_{jC} with the strongest convictions, the interval $[q_{jC}, 1]$ with $q_{jC} = 1 - r_{jC}$, and the open-ended

[12] This uses *rank independence*: a respondent's sustainability rank carries no information about their fee rank, so the distribution of fee ranks is the same at every sustainability rank, and a group defined by its reports on both topics covers a rectangle of sustainability and fee ranks. Appendix B.3 relaxes it: allowing the two convictions to be correlated as strongly as the reports imply changes each effect of the format change by less than a quarter of a percentage point, so the results do not depend on it.

writers are the smaller group at the top, $[q_{jO}, 1]$ with $q_{jO} = 1 - r_{jO}$. Because fewer respondents write about a topic than pick it from the list, $q_{jO} > q_{jC}$: the closed-ended interval reaches further down, to weaker convictions. Writing requires stronger conviction because respondents must retrieve the topic from memory and make the effort of writing it down, while the list only requires them to recognize it (Krosnick 1991, Tourangeau et al. 2000). How much stronger depends on salience: the two thresholds lie further apart for a topic that comes to mind less readily. In the identity surveys of Section 3.2, which rate closeness to two groups on one scale, equally close respondents write about the two groups at very different rates: at the top of the Zurich scale, 16 percent write about cosmopolitan identity and 3 percent about national identity (Appendix E.4.1).

Each topic has four reporting thresholds, one for each question format and position, which let the model fit all four observed reporting shares of the topic. The comparison of the two formats uses the two first-position thresholds, q_{jC} for the closed-ended list and q_{jO} for writing, because neither answer is influenced by an earlier question. The threshold for writing after seeing the list, q_{jP}, is restricted: seeing the list first can make writing easier, but not easier than picking the topic from the list, so it lies between q_{jC} and q_{jO} (Appendix A.2).

The two first-position thresholds divide the conviction ranks into three intervals:

$$U_j = [0, q_{jC}), \qquad K_j = [q_{jC}, q_{jO}), \qquad R_j = [q_{jO}, 1].$$

The three intervals represent: (i) unengaged respondents, who neither write about the topic nor select it from the list; (ii) recognition-only respondents, who choose it from the list but do not write about it; and (iii) recalled respondents, who write about it and also mark it on the list. This assumes that every writer would also select the same topic from the list. In the data, a few respondents write about a topic without choosing it from the list: 1.4–6.5% depending on topic and question order. The estimator therefore matches only the overall shares who select the topic and who write about it, not each respondent's combination of answers. Another version of the model (Appendix B.1.1) lets each respondent write about a topic and select it from the list with probabilities that rise with latent value strength, so a respondent can write about a topic without selecting it.

Each reporting group covers an interval of conviction ranks, and its choices depend on the convictions in that interval. For any interval $[a, b] \subseteq [0, 1]$, the group's rate of value-consistent choice is the average height of the choice curve over the interval,

$$\overline{P}_j[a, b] = \frac{1}{b - a} \int_a^b P_j^{VC}(q)\, dq. \tag{8}$$

The choice rates of the reporting groups are thus informative about the average height of the choice curve over each group's interval, and comparing the groups traces out its shape. Figure 3 draws the three intervals and the groups' rates at the estimates. The estimator also splits the writers by how early the topic first appears in their answer, assuming that earlier mentions reflect stronger convictions (Appendix B.2).

6.4 Two margins of the question format change

Moving from an open to a closed-ended question changes two groups at the same time: target broadening raises the gap, and competing-value screening lowers it (Table 5). The direction of each force does not depend on the functional form. It follows from three assumptions of the model: (i) everyone who writes about a value would also tick it; (ii) the choice moves toward the fund a value favors as its conviction rises; and (iii) the two values' convictions are independent. With the product term, (ii) is not automatic; it holds at the estimates. The functional form determines only the size of each effect.[13] The complete format comparison combines both changes, so its direction depends on which is larger.

Table 5: Two selection margins when moving from open- to closed-ended questions

Margin	Who changes?	Gap effect
Target value	Respondents with weaker target convictions enter the endorser group	Rises
Competing value	Respondents who would recognize the competing value leave the group classified as lacking it	Falls

Note: The table takes the perspective of one target value and its competing value. The size of each movement from one question type to another depends on how many respondents lie between the open and closed-ended intervals for each value and on how consistently respondents act on endorsing a value.

The structural model can switch the question format of the target value and of the competing value separately. Like the OLS regressions of Section 5, the model is estimated on all respondents, and like the OLS-implied gaps, the gap is evaluated for respondents who report the target value but not the competing value. For target value j and competing value k, the gap is

$$G_j(\underbrace{m}_{\text{target}}, \underbrace{n}_{\text{competing}}) = \mathrm{E}[1 - C_j \mid q_j \geq q_{jm},\ q_k < q_{kn}], \qquad m, n \in \{O, C\}, \tag{9}$$

[13] Broadening adds respondents with weaker convictions to the target group; screening removes respondents with stronger competing convictions from the comparison group. By (i) and (iii), each change shifts the distribution of one conviction downward, in the sense of first-order stochastic dominance, and leaves the other unchanged. By (ii), the probability of a value-consistent choice rises with the target conviction and falls with the competing conviction, so broadening lowers the group's average value-consistent choice and screening raises it. Since the gap is one minus this probability, broadening cannot lower the gap, and screening cannot raise it, whichever rule changes first.

where $C_S = Y$ and $C_F = 1 - Y$ denote value-consistent choice, m is the question format that determines who reports the target value, n is the format that determines who counts as not reporting the competing value, and the choice-card distribution is held fixed. The gap is one minus the probability of a value-consistent choice in eq. (6), averaged over the target value's ranks $[q_{jm}, 1]$, the competing value's ranks $[0, q_{kn})$ and the card mix. In the data, the counterpart of eq. (9) is the share of respondents who choose against the target value among those who report it in format m and do not report the competing value in format n:

$$\widehat{G}_j(m,n) = \frac{\#\{i:\ j \text{ reported in } m,\ k \text{ not reported in } n,\ C_{ij}=0\}}{\#\{i:\ j \text{ reported in } m,\ k \text{ not reported in } n\}}, \tag{10}$$

and $\widehat{G}_j(m)$ is the same share without the condition on k. The two mixed gaps, $G_j(C,O)$ and $G_j(O,C)$, have no first-position counterpart in the data: they would require the target value to be reported in one format and the competing value in the other, both as the first question, but each respondent answers only one format first. The model computes them from the first-position thresholds q_{jC} and q_{jO}, and Section 7.2 measures one of them from answers that do not depend on their position. The complete change is $G_j(C,C) - G_j(O,O)$. The model changes each reporting rule separately, but the two changes interact, so the split of the complete change depends on which rule is changed first. The reported effects therefore average the two possible paths (changing the target value's format first or the competing value's format first), which differ by about 0.1 percentage points (Section 7.2):

$$\begin{aligned} T_j &= \frac{1}{2}\Big\{\underbrace{G_j(C,O) - G_j(O,O)}_{\text{competing value open}} + \underbrace{G_j(C,C) - G_j(O,C)}_{\text{competing value closed}}\Big\}, \\ N_j &= \frac{1}{2}\Big\{\underbrace{G_j(O,C) - G_j(O,O)}_{\text{target value open}} + \underbrace{G_j(C,C) - G_j(C,O)}_{\text{target value closed}}\Big\}. \end{aligned} \tag{11}$$

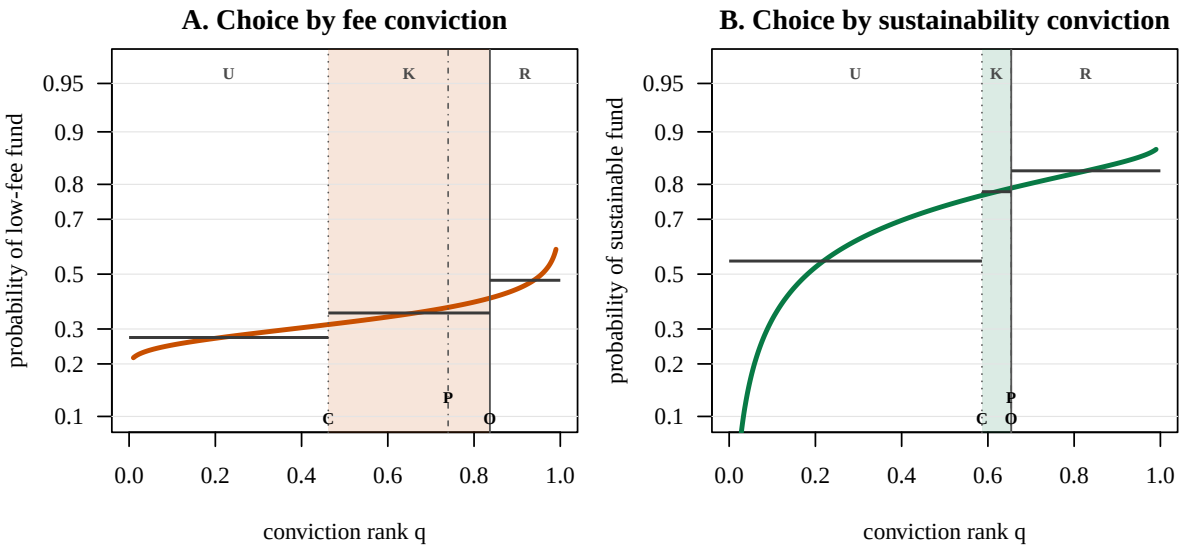
Thus target broadening T_j and competing-value screening N_j, which narrows the comparison group, add to the complete change.

7 Structural Model Results

This section reports the estimates of the model from Section 6. The first part shows how the probability of acting on a value rises with conviction. The second part changes the question format of the target value and of the competing value separately and estimates the two margins of Section 6.4: target broadening, which raises the gap, and competing-value screening, which lowers it. The third part fits the model to three other surveys.

The parameters are chosen so that the model's reporting shares and choice rates match the observed ones as closely as possible (Appendix A).

Figure 3: Structural choice curves



Note: Panels A–B show low-fee and sustainable-fund choice probabilities by conviction rank, holding the competing-value distribution and choice-card mix fixed (eq. (7)). Vertical lines denote closed-ended recognition (C), primed open reporting (P), and unprimed recall (O). The first-position thresholds C and O divide the ranks into the three groups of Section 6.3: unengaged (U), recognition-only (K, shaded) and recalled (R). Each horizontal segment is the group's rate of value-consistent choice in the model, the average of the curve over the group's ranks (eq. (8)). The vertical axis is the log odds, labelled in probabilities; the curves are drawn for ranks 0.01 to 0.99, and the sustainability curve leaves the panel at the lowest ranks.

7.1 From values to action

Between the 10th and the 90th percentile of conviction, the probability of a value-consistent choice rises by

$$\Delta P_S = 51\ [45, 76] \qquad \text{and} \qquad \Delta P_F = 19\ [11, 27]$$

percentage points, from 33 to 85 percent for the sustainable fund and from 25 to 44 percent for the low-fee fund. The brackets give 95% percentile-bootstrap intervals, which re-estimate the full model, including the reporting thresholds, in each bootstrap sample. Sustainability conviction moves the fund choice more than fee conviction. The two scales bend in opposite directions: the sustainability skew is -0.76, so its index is bounded above and the strongest sustainability convictions add little, while the fee skew is 0.33, which stretches the top of the fee ranking. Only the sustainability skew differs significantly from zero; its size depends on how respondents who write without ticking are counted

(Appendix B.4). The product coefficient, 0.05, does not differ from zero (Appendix Table A.1).

Figure 3 shows the whole curves: the probability of making a value-consistent choice at each conviction quantile, selecting the low-fee fund for fee conviction and the sustainable fund for sustainability conviction. On the log-odds scale of the figure, the sustainability curve flattens among strong convictions and the fee curve steepens, as the three groups of Table 4 do: estimated by logit instead of OLS, the ticked-only group lies 0.67 of the way from the non-reporters to the writers for sustainability, but 0.59 for fees (column 5; 0.64 and 0.42 in column 6). In this survey, the contrast is concentrated on the rating cards, which show the funds' sustainability rating: on the control card, the ticked-only group lies at similar positions for the two values (0.38 for sustainability and 0.42 for fees, from the raw group rates), and the model's sustainability skew is close to zero there (Appendix B.4). The model does not identify why the curves bend. The contrast follows the two values' salience, but the rating may also flatten the sustainability curve: it gives respondents with a mild concern for sustainability a reason to choose the sustainable fund. The three other surveys, which have no rating cards, show the same contrast between the more and the less salient value in the estimates (Appendix Figure E.2), although the skews are jointly significant only in the Zurich survey (Appendix B.4).

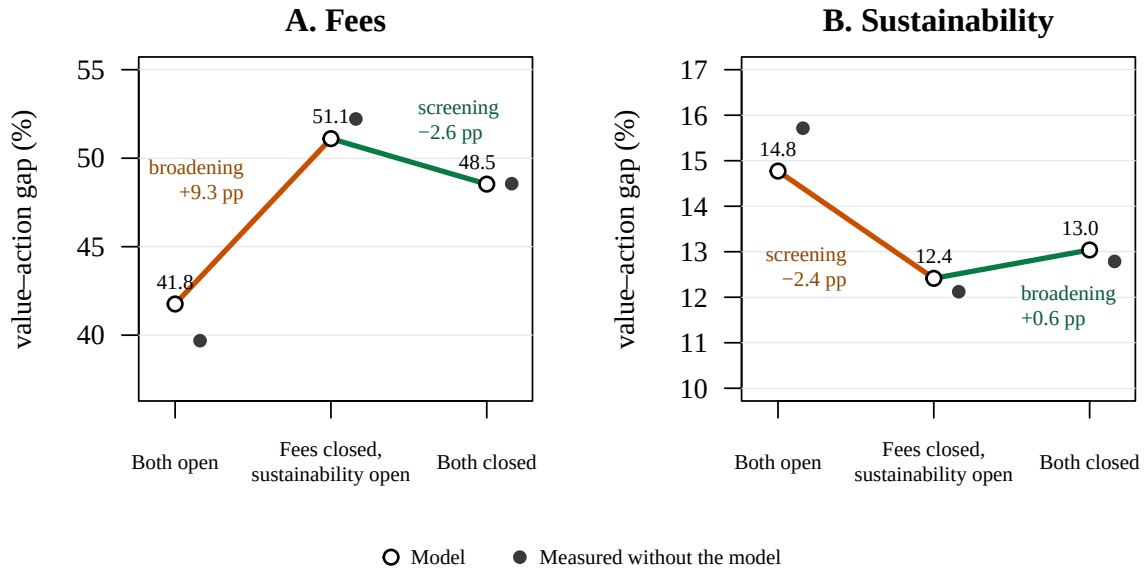
Fee gaps exceed sustainability gaps in both formats because the two values point to opposite funds and sustainability conviction moves the choice more than fee conviction: most investors who report both values give up low fees.

The two question formats capture different parts of each conviction ranking. In the data, the recognition-only interval, between the closed-ended and the open-ended threshold, covers 37 percent of respondents for fees (54 percent select fees from the list, 17 percent write about them) but only 6 percent for sustainability (41 and 35 percent). In the model, 27, 35, and 48 percent of unengaged, recognition-only, and recalled respondents choose the low-fee fund; the corresponding rates for the sustainable fund are 55, 78, and 83 percent. Thus closed-ended responses add a large group of fee respondents who act on fees less often than fee writers, which raises the fee gap. Yet this group acts on fees more often than the unengaged. Removing the group from the respondents who do not report fees therefore lowers the sustainability gap: sustainability reporters who leave fees unticked are rarely pulled toward the low-fee fund. For sustainability, the added group is small. The sizes of the two margins depend on the latent scale: on a log-rank scale, fee broadening is 3.3 instead of 9.4 points. The same margin dominates for each value under every scale and under a hurdle version of the choice, in which a value enters the choice only with a probability that rises with conviction (Appendices B.4 and B.5).

7.2 Moving from open to closed-ended questions

Section 6.4 defined the two margins; this subsection estimates them for the randomized comparison of first answers. Figure 4 changes the fee format first and the sustainability format second, the path that the data can also measure; the other path in eq. (11) gives a split that differs by about 0.1 percentage points, and the text reports the average of the two paths.

Figure 4: Gaps when changing one reporting rule at a time, the fee format first: model and data



Note: Each panel begins with the first-position gap when both values are reported in the open format and changes the fee format first and the sustainability format second (eq. (12)); the text averages this path and the reverse one, which the data cannot measure. For fees, the first segment is broadening, which adds the recognition-only fee respondents K_F to the fee writers R_F, and the second is the screen by sustainability, which removes K_S from the group that does not report sustainability, $U_S \cup K_S$; for sustainability, the screen by fees comes first and broadening second. Each segment has the color of the value whose format changes (orange fees, green sustainability). Circles: the model; dots: the gaps measured without the model (eq. (10)), the middle one from both question orders (Appendix B.8). Both count a value written but not ticked as reported in both formats.

For fees, broadening the target group increases the gap by 9.4 percentage points, and the stricter screen of the competing sustainability value offsets 2.6 points; the net change is +6.8 points. For sustainability, broadening adds only 0.7 points, and the stricter screen of the competing fee value lowers the gap by 2.4 points; the net change is −1.7 points. Broadening dominates for fees in 99 percent of the bootstrap samples. For sustainability, screening dominates in nine of ten samples, and the interval of the net change includes zero (Appendix Table B.4): the split is clear for fees and less certain for sustainability. In the version of the model in which respondents can write about a topic without ticking it (Section 6.3), the same margin dominates for each value.

The same fee-reporting asymmetry drives both results: many respondents recognize fees on the list, but few mention them without prompting. When fees are the target,

closed-ended recognition adds the large group that recognizes fees but does not write about them, raising the fee gap. When fees are the competing value, that group sits among the respondents who do not write about fees, and its hidden fee concern pushes some sustainability writers toward the low-fee fund. Closed-ended nonselection excludes these respondents and lowers the sustainability gap.

The data support this split. Because each respondent answers one format first, the randomized comparison observes $\widehat{G}_j(O,O)$ in the open-first group and $\widehat{G}_j(C,C)$ in the closed-first group (eq. (10)). The mixed gap with fees closed and sustainability open is observed in both groups, from answers that do not depend on their position (Appendix B.8). The path through this gap changes the fee format first:

$$\begin{aligned} \widehat{T}_F &= \widehat{G}_F(C,O) - \widehat{G}_F(O,O), & \widehat{N}_F &= \widehat{G}_F(C,C) - \widehat{G}_F(C,O), \\ \widehat{N}_S &= \widehat{G}_S(O,C) - \widehat{G}_S(O,O), & \widehat{T}_S &= \widehat{G}_S(C,C) - \widehat{G}_S(O,C). \end{aligned} \tag{12}$$

Measured this way, broadening again dominates for fees and screening for sustainability (dots in Figure 4). The model reproduces the direction of both raw changes with smaller estimates, within sampling error, and a simpler version, without card, earliness or priming components, fitted to the open-first group, predicts the first-position gaps of the closed-first group within 4 points, whereas assuming no format effect misses the fee gaps by up to 14 points; fitted the other way round, it misses the open-first fee gaps by 7 and 8 points (Appendix B.8).

7.3 Evidence from other surveys

The results so far come from one survey, in which fees compete with sustainability in a fund choice. With three other surveys, I check whether the two margins, the weak convictions that a list admits and the clearer "no" of an unticked box, also appear for other values and outcomes (Section 3.2). Because their respondents always write before they tick, this evidence corresponds to this survey's paired answers (Appendix B.8) rather than to its randomized comparison. In Andre et al. (2022), households forecast inflation, and a cost channel, which implies rising inflation, competes with a demand channel, which implies falling inflation. The respondents I use read one of three scenarios, in which the oil price, the interest rate or income taxes rise, and ticked channels on a list for that scenario; the estimates below average the three scenarios. In two surveys on political identities, national identity pulls the vote away from the left, and cosmopolitan identity toward it.

In this survey, net of the other value, the list raises the fee gap and lowers the sustainability gap. In the model, both changes come mainly from the many respondents who tick fees without writing about them, the group that makes fees the less salient

value. The question is whether, in the other surveys, the list again raises the gap of the less salient value and lowers the gap of the more salient value, where the less salient value is the one to which the list adds more respondents per writer.

Survey data from Andre et al. (2022) show the same two margins. Both the cost and the demand channel are written by few respondents (13 and 16 percent) and ticked by many (58 and 56 percent). In the raw data, the gap, the share of a channel's reporters whose forecast does not follow the channel, rises from written to ticked reports for the cost channel, which most forecasts follow, by 7.2 percentage points net of the demand channel, as for fees; for the demand channel, it falls by 8.6 points net of the cost channel, as for sustainability. Both changes are significant at the 5% level (Appendix Table E.4).

Fitted to these data, with the forecast in place of the fund choice and the two convictions allowed to correlate, the model splits each change into its two margins. With three possible forecasts, cost conviction raises the utility of a rising forecast and demand conviction that of a falling one, and holding both makes either direction less likely than no change (Appendix E.3). For the cost channel, broadening raises the gap by 7.5 points, and the screen lowers it by only 1.3. For the demand channel, the stricter screen of the cost channel lowers the gap by 12.3 points, and broadening raises it by 3.7. The model reproduces both raw changes with smaller estimates when they are counted as in the model, with respondents who write a channel without ticking it among its reporters in both formats (Appendix E.3). The two screens differ although the list adds similar shares of respondents to both channels: in each scenario, cost conviction moves the forecast more than demand conviction, so excluding respondents who tick the cost channel lowers the demand gap more than the reverse. Because both answers explain a forecast already made, they may partly rationalize it (Appendix E.3).

In the two surveys on political identities, the Zurich survey and the Swiss survey of 2018, national identity is, like fees, rarely written but widely endorsed on the closeness scale (by 2.3 and 72.4 percent in the Zurich survey). Its gap is the share of its reporters who vote for the left; here the gap measures how often the vote does not follow the reported identity, which need not be an inconsistency. As with the fee gap net of sustainability, the list raises the gap of national identity net of cosmopolitan identity: by 5.2 points in the Zurich survey and by 8.2 points in the Swiss survey, although both are imprecisely estimated. As with the sustainability gap net of fees, it lowers the gap of cosmopolitan identity net of national identity: by 8.3 points in the Zurich survey and, less precisely, by 7.7 points in the Swiss survey (Appendix Table E.6).

Figure 5 shows, for each value, the change in the gap from written to closed-ended reports, one value at a time and net of the other value (eq. (10)):

$$\widehat{G}_j(C) - \widehat{G}_j(O) \qquad \text{and} \qquad \widehat{G}_j(C,C) - \widehat{G}_j(O,O).$$

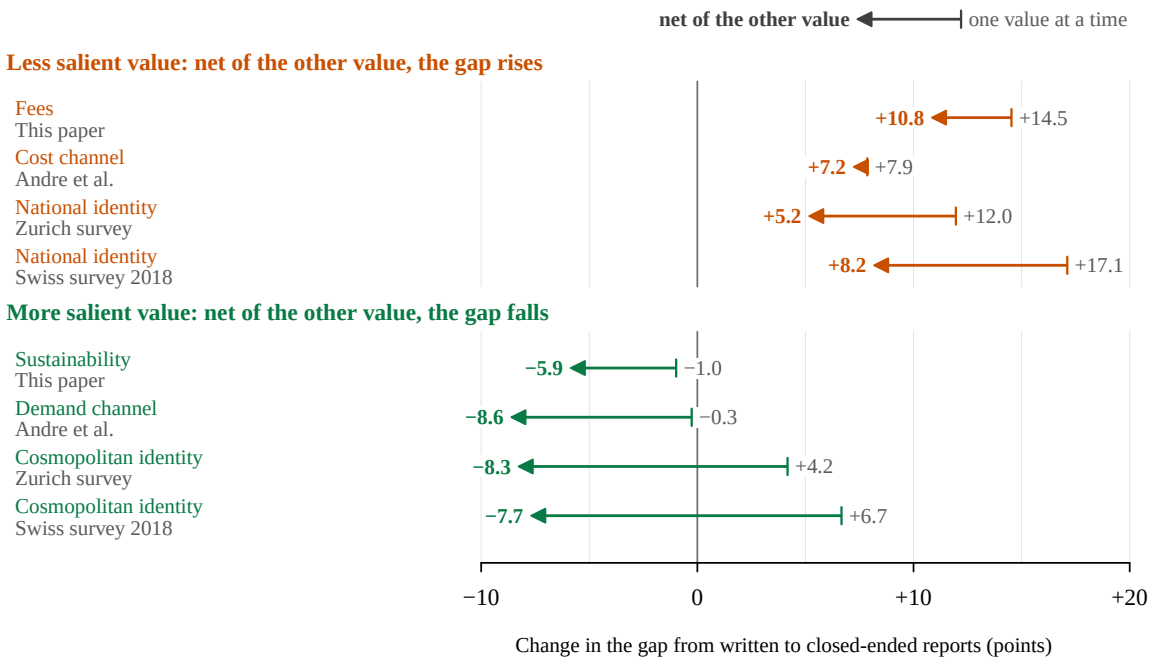
In all four surveys, including this one, these estimates, averaged over scenarios and countries, show the same pattern: net of the other value, the list raises the gap of the less salient value and lowers the gap of the more salient value. Measured one value at a time, the gap of the less salient value rises more than that of the more salient value in each of the four surveys, even in Andre et al. (2022), where the two channels differ little in salience. In each of the three other surveys, netting out the other value lowers the change more for the more salient value: its arrow in Figure 5 is longer.

Because these surveys compare each respondent's own open and closed answers, the two margins can also be measured without the model, by computing eq. (11) from the four gaps $\widehat{G}_j(m,n)$ (dots in Figure 6; Appendix Table E.9), and so can those of this survey's randomized comparison, along the path of eq. (12). Measured this way, broadening raises the gap and screening lowers it for each value in the four surveys, although several intervals include zero. Fitted to the same answers, with the same specification in all four surveys, the model comes close to the measured broadening, except for fees in this survey, the cost channel in Andre et al. (2022) and cosmopolitan identity in the Zurich survey, where it understates it, and its screens lie within 2 points of the measured ones. The product term matters only in the survey of Andre et al. (2022), with its three possible forecasts (Appendix B.4). The other surveys also estimate a correlation between the two convictions; this survey's reports put it close to zero (Appendix B.3). As an out-of-sample check, I fit the simpler version of the model from Section 7.2 to three countries of the Zurich survey at a time and predict the margins of the fourth country. These predictions miss the measured margins about as much as predictions by the average margins measured in the other three countries; the hurdle version of the model predicts them better (Appendix E.4.2). In this check and in the question-order check of Section 7.2, the model with normal scores predicts as well as the skewed model or slightly better, within sampling error (Appendices E.4.2 and E.4.3).

In the model, the screen accounts for most of the difference between the two values' gap changes in two of the three other surveys and for about half of it in the Zurich survey: the competing value's screen lowers the gap of the more salient value two to ten times as much as that of the less salient value (Figure 6, outlined bars). The measured broadening (dots) differs between the two values about as much as the model's, more in Andre et al. (2022) and less in the Zurich survey; the screen accounts for more than half of the measured difference in these surveys. In this survey, the model's broadening carries

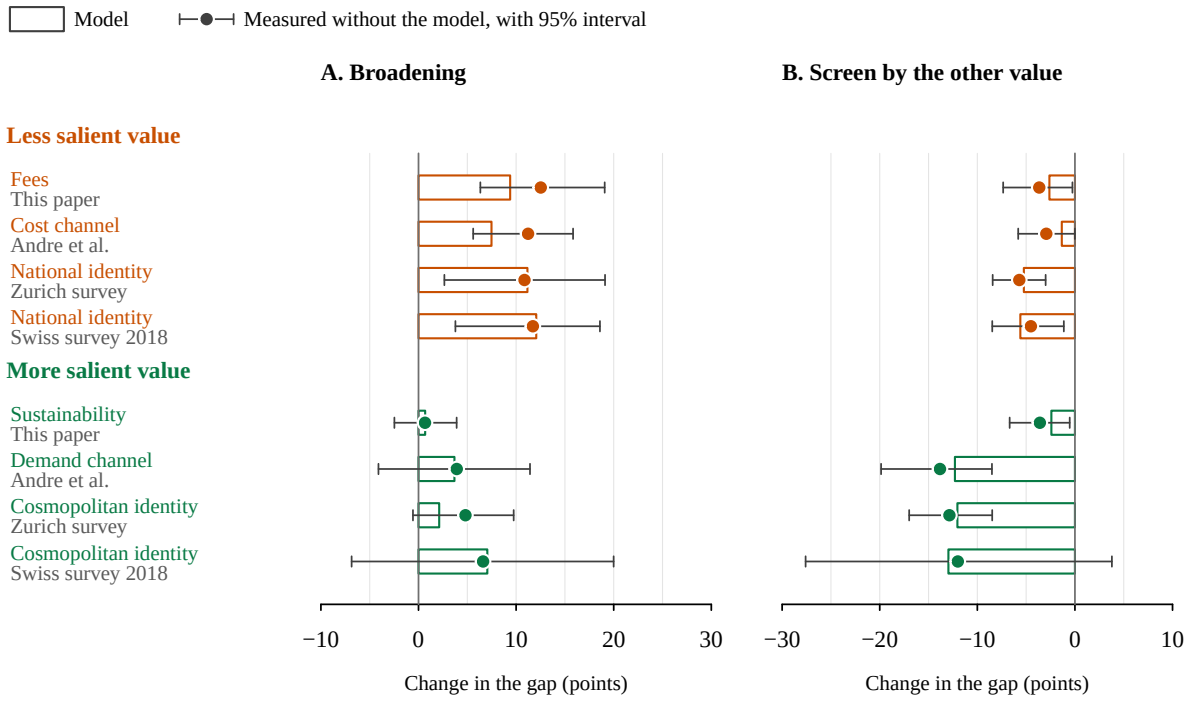
most of the difference, as the measured broadening does: 9.4 points for fees against 0.7 for sustainability in the first answers.

Figure 5: Question format and the value–action gap across surveys



Note: Each arrow starts at the change in the gap from written to closed-ended reports one value at a time (grey number) and ends at the raw change net of the other value of the survey, as in Section 5.1 (bold number). In each survey, orange marks the less salient value (the one for which the list adds more respondents per writer; Table 2) and green the more salient value. Closed-ended reports are ticks on a list in this paper and in Andre et al. (2022), a closeness of at least 7 on a scale from 1 to 10 in the Zurich survey, and "rather" or "very close" in the Swiss survey of 2018. For this paper, the changes compare the randomized question orders. For the other surveys, they compare each respondent's two answers within each scenario or country, averaged with weights proportional to its number of respondents, so that differences between scenarios or countries do not enter the change. A value written but not reported in the closed-ended question counts as not reported there, as in Section 5. Appendix Tables E.4 and E.6 report standard errors.

Figure 6: The two margins across surveys: model and data



Note: Outlined bars show the model's two margins for each value (this paper: Section 7.2; the other surveys: Appendix Tables E.5 and E.7); dots show the same margins measured without the model, and lines their 95% intervals (Appendix Table E.9, which also reports the model's standard errors). For the other surveys, the dots use the paired answers; for this paper, they follow the path of eq. (12) in the randomized comparison of first answers. Bars and dots count a value written but not reported in the closed-ended question as reported in both formats, as the model does, so this paper's measured margins add up to 8.9 and −2.9 points rather than the 10.8 and −5.9 of Figure 5. Broadening (Panel A) is the change in the gap when the value itself moves from written to ticked reports; the screen by the other value (Panel B) is the change when the other value counts as unreported only if it is neither written nor ticked, negative when the gap falls. For the other surveys, the margins are averaged over scenarios and countries.

8 Conclusion

This paper revisits a long-standing question in survey design: when do open- or closed-ended questions provide value measures that align best with subsequent choices? I argue that question format acts as a selection device for latent value strength. Because expressing a value requires attention and effort, closed-ended questions lower the cost of expression and admit weakly held values, whereas open-ended questions raise that cost and select stronger convictions; at the same time, an unticked box is a clearer "no" than silence in a text field. Using the value–action gap as a benchmark for alignment, I show that, for a given topic, the gap changes when the question switches from open to closed-ended, and that the size and even the direction of this change differ between topics.

In a randomized survey of 3,059 Swiss retail investors choosing between a hypothetical high-fee sustainable fund and a less sustainable low-fee fund, this mechanism explains the different prevalence and alignment of stated attitudes under the two question formats. Measured one value at a time, moving from open-ended to closed-ended elicitation in first position raises the fee value–action gap from 48.1% to 62.6%, while the sustainability gap moves only from 17.0% to 16.0%: for sustainability, the more salient attribute in this survey, closed-ended endorsements align with the sustainable choice about as well as open-ended answers, whereas for the less salient fees the closed-ended question admits many weak statements of fee importance that translate less consistently into choice. With the competing value added, the fee gap rises from 39.7% to 50.5%, while the sustainability gap falls from 15.7% to 9.9%. Three other surveys with paired answers, on inflation expectations and political identities, show the same two effects in their estimates: net of the other value, the list raises the gap of the less salient value and lowers that of the more salient value.

I formalize this mechanism with a micro-founded structural model in which the survey answers are thresholded readings of a latent conviction and the fund choice weighs each value by its conviction. Comparing the two formats under a common choice rule, the model attributes the rise in the fee gap to target-value broadening, net of a stricter sustainability screen, and the fall in the sustainability gap to a stricter fee screen, net of broadening; the sustainability split is less certain than the fee split. The model thus reproduces the direction of both raw changes with smaller estimates, but the differences are within sampling error. Fitted to the three other surveys with the same specification, the model comes close to most of the measured margins. A single choice can hardly separate a value that does not enter a decision from one that enters with a low weight (for repeated choices, see Hess et al. 2013), but choices at several fee differences could.

The fee–sustainability contrast shows why both target and competing values matter. Both changes stem mainly from one group: the many respondents who recognize fees on the list but do not write about them. When fees are the target, the closed-ended question adds this group, which acts on fee concern less consistently than fee writers, and raises the fee gap. When fees are the competing value, the same group is hidden among respondents who do not write about fees; closed-ended non-selection excludes it, which creates a cleaner comparison group and lowers the sustainability gap. The competing value also explains why fee statements align less well with choices than sustainability statements in both formats: sustainability conviction moves the choice more than fee conviction, and most investors who report both values give up low fees.

These findings yield a practical rule for elicitation that adds a behavioral criterion to existing guidance on when to use each format (Haaland et al. 2025). Open-ended answers are selective positive signals; closed-ended non-selections are selective negative signals. The format should therefore be chosen with both the target value and its competitor in mind. For the target value, closed-ended answers align with choices about as well as open-ended answers when the list adds few respondents beyond those who write, as for sustainability here; when the list adds many respondents who recognize a value without recalling it, closed-ended answers pool strong and weak convictions and align less well with choices. For the competing value, closed-ended non-selection is most informative when the value is less salient, as fees are here, and carries weight in the choice. The other surveys suggest that the rule extends beyond investing: for inflation forecasts and votes, too, the list adds endorsers who act on a value less consistently, and an unticked competing value says more than silence in an open answer. When both formats are used, asking the open-ended question first, as Andre et al. (2022) and Andre et al. (2026) do, keeps written answers unprimed; in this survey, seeing the list first raised written fee mentions from 17% to 26%.

Beyond survey design, the results speak to policies that rely on stated values to steer decisions. For MiFID-II, which requires advisers to elicit clients' sustainability priorities, often with tick boxes, this survey offers little support for the concern that closed-ended questions artificially over-report them: the list added few sustainability endorsements, and closed-ended endorsers chose the sustainable fund about as often as respondents who wrote about sustainability. When policymakers seek to increase fee sensitivity, interventions that increase the attitude salience of fees may help: for example, by supporting fee concerns through education, personalized long-run cost framing, and repeated feedback about fee drag. Their effect is limited where fees compete with sustainability, however: most investors who report both values choose the sustainable fund. These results concern choices made after both questions, with both attributes on the choice card; whether val-

ues that a list brings to mind also shape decisions in which nothing presents the attribute remains open.

References

Ajzen, I.: 1991, The theory of planned behavior, *Organizational behavior and human decision processes* **50**(2), 179–211.

Ajzen, I. and Fishbein, M.: 1973, Attitudinal and normative variables as predictors of specific behavior., *Journal of personality and Social Psychology* **27**(1), 41–57.

Akerlof, G. A. and Kranton, R. E.: 2000, Economics and identity, *The quarterly journal of economics* **115**(3), 715–753.

Alvarez, R. M. and Nagler, J.: 1998, When politics and models collide: Estimating models of multiparty elections, *American Journal of Political Science* **42**(1), 55–96.

Anderson, A. and Robinson, D. T.: 2022, Financial literacy in the age of green investment, *Review of Finance* **26**(6), 1551–1584.

Andre, P., Haaland, I., Roth, C., Wiederholt, M. and Wohlfart, J.: 2026, Narratives about the macroeconomy, *The Review of Economic Studies*, advance access. doi: 10.1093/restud/rdag014.

Andre, P., Pizzinelli, C., Roth, C. and Wohlfart, J.: 2022, Subjective Models of the Macroeconomy: Evidence From Experts and Representative Samples, *The Review of Economic Studies* **89**(6), 2958–2991.

Barber, B. M., Odean, T. and Zheng, L.: 2005, Out of sight, out of mind: The effects of expenses on mutual fund flows, *The Journal of Business* **78**(6), 2095–2120.

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

Behr, D., Bandilla, W., Kaczmirek, L. and Braun, M.: 2014, Cognitive probes in web surveys: On the effect of different text box size and probing exposure on response quality, *Social Science Computer Review* **32**(4), 524–533.

Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Boersch-Supan, A., Brownstone, D., Bunch, D. S., Daly, A., de Palma, A., Gopinath, D., Karlstrom, A. and Munizaga, M. A.: 2002, Hybrid choice models: Progress and challenges, *Marketing Letters* **13**(3), 163–175.

Bertrand, M., Kamenica, E. and Pan, J.: 2015, Gender identity and relative income within households, *The Quarterly Journal of Economics* **130**(2), 571–614.

Bertrand, M. and Mullainathan, S.: 2001, Do people mean what they say? implications for subjective survey data, *American Economic Review* **91**(2), 67–72.

Bethlendi, A., Nagy, L. and Póra, A.: 2022, Green finance: the neglected consumer demand, *Journal of Sustainable Finance & Investment* pp. 1–19.

Blake, J.: 1999, Overcoming the 'value-action gap' in environmental policy: Tensions between national policy and local experience, *Local Environment* **4**(3), 257–278.

Bornschier, S., Haffert, L., Häusermann, S., Steenbergen, M. and Zollinger, D.: 2024, *Cleavage Formation in the 21st Century*, Elements in European Politics, Cambridge University Press, Cambridge.

Bornschier, S., Häusermann, S., Zollinger, D. and Colombo, C.: 2021, How "us" and "them" relates to voting behavior: Social structure, social identities, and electoral choice, *Comparative Political Studies* **54**(12), 2087–2122.

Chinco, A., Hartzmark, S. M. and Sussman, A. B.: 2022, A new test of risk factor relevance, *The Journal of Finance* **77**(4), 2183–2238.

Choi, J. J., Laibson, D. and Madrian, B. C.: 2010, Why does the law of one price fail? an experiment on index mutual funds, *The Review of Financial Studies* **23**(4), 1405–1432.

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

Cragg, J. G.: 1971, Some statistical models for limited dependent variables with application to the demand for durable goods, *Econometrica* **39**(5), 829–844.

Daly, A., Hess, S., Patruni, B., Potoglou, D. and Rohr, C.: 2012, Using ordered attitudinal indicators in a latent variable choice model: A study of the impact of security on rail travel behaviour, *Transportation* **39**(2), 267–297.

Dechezleprêtre, A., Fabre, A., Kruse, T., Planterose, B., Sanchez Chico, A. and Stantcheva, S.: 2025, Fighting climate change: International attitudes toward climate policies, *American Economic Review* **115**(4), 1258–1300.

Engler, D., Gutsche, G. and Ziegler, A.: 2025, Does the willingness to pay for sustainable investments differ between non-incentivized and incentivized choice experiments?, *Technical Report 15-2025*, MAGKS Joint Discussion Paper Series in Economics, Philipps-University Marburg, School of Business and Economics.

Enke, B. and Graeber, T.: 2023, Cognitive uncertainty, *The Quarterly Journal of Economics* **138**(4), 2021–2067.

European Commission: 2021, Commission delegated regulation (eu) 2021/1253 of 21 april 2021 amending delegated regulation (eu) 2017/565 as regards the integration of

sustainability factors, risks and preferences into certain organisational requirements and operating conditions for investment firms, *Official Journal of the European Union*, L 277/1.
URL: *https://eur-lex.europa.eu/eli/reg˙del/2021/1253/oj*

Falk, A., Becker, A., Dohmen, T., Huffman, D. and Sunde, U.: 2023, The preference survey module: A validated instrument for measuring risk, time, and social preferences, *Management Science* **69**(4), 1935–1950.

Fazio, R. H. and Williams, C. J.: 1986, Attitude accessibility as a moderator of the attitude-perception and attitude-behavior relations: An investigation of the 1984 presidential election, *Journal of Personality and Social Psychology* **51**(3), 505–514.

Fehr, E., Epper, T. and Senn, J.: 2024, Social preferences and redistributive politics, *Review of Economics and Statistics* pp. 1–45.

Ferrario, B. and Stantcheva, S.: 2022, Eliciting people's first-order concerns: Text analysis of open-ended survey questions, *AEA Papers and Proceedings* **112**, 163–69.

Fifer, S., Rose, J. and Greaves, S.: 2014, Hypothetical bias in stated choice experiments: Is it a problem? and if so, how do we deal with it?, *Transportation research part A: policy and practice* **61**, 164–177.

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

Funk, P.: 2016, How accurate are surveyed preferences for public policies? evidence from a unique institutional setup, *Review of Economics and Statistics* **98**(3), 442–454.

Geer, J. G.: 1988, What do open-ended questions measure?, *Public Opinion Quarterly* **52**(3), 365–367.

Geer, J. G.: 1991, Do open-ended questions measure "salient" issues?, *Public Opinion Quarterly* **55**(3), 360–370.

Gennaioli, N. and Shleifer, A.: 2010, What comes to mind, *The Quarterly Journal of Economics* **125**(4), 1399–1433.

Gibson Brandon, R., Sohn, M., Tanner, C. and Wagner, A. F.: 2023, Earnings management and the role of moral values in investing, *European Accounting Review* pp. 1–31.

Gillen, B., Snowberg, E. and Yariv, L.: 2019, Experimenting with measurement error: Techniques with applications to the caltech cohort study, *Journal of Political Economy* **127**(4), 1826–1863.

Grave, E., Bojanowski, P., Gupta, P., Joulin, A. and Mikolov, T.: 2018, Learning word vectors for 157 languages, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Grossman, G. M. and Helpman, E.: 2021, Identity politics and trade policy, *The Review of Economic Studies* **88**(3), 1101–1126.

Haaland, I., Roth, C., Stantcheva, S. and Wohlfart, J.: 2025, Understanding economic behavior using open-ended survey data, *Journal of Economic Literature* **63**(4), 1244–1280.

Hadler, P.: 2025, The effects of open-ended probes on closed survey questions in web surveys, *Sociological Methods & Research* **54**(1), 106–139.

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

Hainmueller, J., Hangartner, D. and Yamamoto, T.: 2015, Validating vignette and conjoint survey experiments against real-world behavior, *Proceedings of the National Academy of Sciences* **112**(8), 2395–2400.

Handel, B. R. and Kolstad, J. T.: 2015, Health insurance for "humans": Information frictions, plan choice, and consumer welfare, *American Economic Review* **105**(8), 2449–2500.

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

Heeb, F., Kölbel, J. F., Paetzold, F. and Zeisberger, S.: 2023, Do investors care about impact?, *The Review of Financial Studies* **36**(5), 1737–1787.

Hess, S. and Hensher, D. A.: 2013, Making use of respondent reported processing information to understand attribute importance: A latent variable scaling approach, *Transportation* **40**(2), 397–412.

Hess, S., Stathopoulos, A., Campbell, D., O'Neill, V. and Caussade, S.: 2013, It's not that I don't care, I just don't care very much: confounding between attribute non-attendance and taste heterogeneity, *Transportation* **40**(3), 583–607.

Hoaglin, D. C.: 1985, Summarizing shape numerically: The *g*-and-*h* distributions, *in* D. C. Hoaglin, F. Mosteller and J. W. Tukey (eds), *Exploring Data Tables, Trends, and Shapes*, Wiley, New York, pp. 461–513.

Hole, A. R.: 2011, A discrete choice model with endogenous attribute attendance, *Economics Letters* **110**(3), 203–205.

Hole, A. R., Kolstad, J. R. and Gyrd-Hansen, D.: 2013, Inferred vs. stated attribute non-attendance in choice experiments: A study of doctors' prescription behaviour, *Journal of Economic Behavior & Organization* **96**, 21–31.

Honnibal, M., Montani, I., Van Landeghem, S. and Boyd, A.: 2020, spaCy: Industrial-strength Natural Language Processing in Python, Zenodo.

Imbens, G. W. and Manski, C. F.: 2004, Confidence intervals for partially identified parameters, *Econometrica* **72**(6), 1845–1857.

Kollmuss, A. and Agyeman, J.: 2002, Mind the gap: Why do people act environmentally and what are the barriers to pro-environmental behavior?, *Environmental Education Research* **8**(3), 239–260.

Kraft, P. W.: 2024, Women also know stuff: Challenging the gender gap in political sophistication, *American Political Science Review* **118**(2), 903–921.

Krosnick, J. A.: 1990, Government policy and citizen passion: A study of issue publics in contemporary america, *Political behavior* **12**(1), 59–92.

Krosnick, J. A.: 1991, Response strategies for coping with the cognitive demands of attitude measures in surveys, *Applied Cognitive Psychology* **5**(3), 213–236.

Krosnick, J. A.: 1999, Survey research, *Annual Review of Psychology* **50**(1), 537–567.

Laurent, G., Kapferer, J.-N. and Roussel, F.: 1995, The underlying structure of brand awareness scores, *Marketing Science* **14**(3), G170–G179.

Lazarsfeld, P. F.: 1944, The controversy over detailed interviews—an offer for negotiation, *Public opinion quarterly* **8**(1), 38–60.

Lee, D. S.: 2009, Training, wages, and sample selection: Estimating sharp bounds on treatment effects, *The Review of Economic Studies* **76**(3), 1071–1102.

Lusardi, A. and Mitchell, O. S.: 2008, Planning and financial literacy: How do women fare?, *American Economic Review* **98**(2), 413–17.

Manzini, P. and Mariotti, M.: 2014, Stochastic choice and consideration sets, *Econometrica* **82**(3), 1153–1176.

Masatlioglu, Y., Nakajima, D. and Ozbay, E. Y.: 2012, Revealed attention, *American Economic Review* **102**(5), 2183–2205.

McFadden, D.: 1974, Conditional logit analysis of qualitative choice behavior, *in* P. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, pp. 105–142.

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

Meitinger, K. and Behr, D.: 2016, Comparing cognitive interviewing and online probing: Do they find similar results?, *Field Methods* **28**(4), 363–380.

Morwitz, V. G., Johnson, E. and Schmittlein, D.: 1993, Does measuring intent change behavior?, *Journal of Consumer Research* **20**(1), 46–61.

Nguyen, H. V., Nguyen, C. H. and Hoang, T. T. B.: 2019, Green consumption: Closing the intention-behavior gap, *Sustainable Development* **27**(1), 118–129.

Nimon, W. and Beghin, J.: 1999, Are eco-labels valuable? evidence from the apparel industry, *American Journal of Agricultural Economics* **81**(4), 801–811.

Raghunandan, A. and Rajgopal, S.: 2022, Do esg funds make stakeholder-friendly investments?, *Review of Accounting Studies* **27**(3), 822–863.

Reja, U., Manfreda, K. L., Hlebec, V. and Vehovar, V.: 2003, Open-ended vs. close-ended questions in web questionnaires, *Developments in Applied Statistics* **19**(1), 159–177.

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

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B. and Rand, D. G.: 2014, Structural topic models for open-ended survey responses, *American Journal of Political Science* **58**(4), 1064–1082.

Roheim, C. A., Asche, F. and Santos, J. I.: 2011, The elusive price premium for eco-labelled products: evidence from seafood in the uk market, *Journal of Agricultural Economics* **62**(3), 655–668.

Schuman, H., Ludwig, J. and Krosnick, J. A.: 1986, The perceived threat of nuclear war, salience, and open questions, *Public Opinion Quarterly* **50**(4), 519–536.

Schuman, H. and Presser, S.: 1979, The open and closed question, *American Sociological Review* **44**(5), 692–712.

Schuman, H. and Scott, J.: 1987, Problems in the use of survey questions to measure public opinion, *Science* **236**(4804), 957–959.

Singer, E. and Couper, M. P.: 2017, Some methodological uses of responses to open questions and other verbatim comments in quantitative surveys, *Methods, data, analyses: a journal for quantitative methods and survey methodology (mda)* **11**(2), 115–134.

Sprott, D. E., Spangenberg, E. R., Block, L. G., Fitzsimons, G. J., Morwitz, V. G. and Williams, P.: 2006, The question–behavior effect: What we know and where we go from here, *Social Influence* **1**(2), 128–137.

Strachan, D. and Gupta, I.: 2022, Sustainability preferences: Complex new eu rules require collection of clients' esg preferences from august 2022, *Deloitte UK — Financial Services blog*.
URL: *https://www.deloitte.com/uk/en/Industries/financial-services/blogs/sustainability-preferences-complex-new-eu-rules-require-collection-of-clients-esg-preferences-from-august-2022.html*

Tourangeau, R., Rips, L. J. and Rasinski, K.: 2000, *The Psychology of Survey Response*, Cambridge University Press, Cambridge.

Vermeir, I. and Verbeke, W.: 2006, Sustainable food consumption: Exploring the consumer "attitude–behavioral intention" gap, *Journal of Agricultural and Environmental ethics* **19**, 169–194.

Verplanken, B. and Holland, R. W.: 2002, Motivated decision making: Effects of activation and self-centrality of values on choices and behavior, *Journal of Personality and Social Psychology* **82**(3), 434–447.

Vij, A. and Walker, J. L.: 2016, How, when and why integrated choice and latent variable models are latently useful, *Transportation Research Part B: Methodological* **90**, 192–217.

Walker, J. L.: 2001, *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables*, PhD thesis, Massachusetts Institute of Technology.

Wekhof, T. and Houde, S.: 2023, Using narratives to infer preferences in understanding the energy efficiency gap, *Nature Energy* **8**(9), 965–977.

White, K., Hardisty, D. J. and Habib, R.: 2019, The elusive green consumer, *Harvard Business Review* **97**(4), 124–133.

Zollinger, D.: 2024, Cleavage identities in voters' own words: Harnessing open-ended survey responses, *American Journal of Political Science* **68**(1), 139–159.

Zollinger, D. and Häusermann, S.: 2026, Status grievances and the far right vote, *European Journal of Political Research* . Forthcoming.

A Choice Probabilities and Estimation

This appendix provides the derivation and estimation details for the structural model. It explains how the two convictions translate into the choice probabilities of a reporting group, how the parameters are estimated, and how averaging over the competing value and the randomized cards produces the choice curves shown in the main text.

A.1 Choice probabilities of a reporting group

Section 6.2 gives the probability of choosing the sustainable fund at given convictions, eq. (6). Let $Y_i = 1$ denote choosing the sustainable, high-fee fund. Let $D_i = 0$ denote the control card and $D_i = 1$ any of the three ESG-rating cards, which are pooled in estimation. The data do not show convictions, only the reports, and a group defined by its reports covers a range of conviction ranks: $B_{j,c} = [\underline{q}_{j,c}, \overline{q}_{j,c}]$ for value $j \in \{S, F\}$ in group c (Appendix A.2). Because each rank is uniform and the two ranks are, by assumption, independent (relaxed in Appendix B.3), the group's ranks are uniform on the rectangle $B_{S,c} \times B_{F,c}$, and its probability of choosing the sustainable fund on card d averages the logit over the rectangle:

$$p_c(\theta) = \frac{1}{(\overline{q}_{S,c} - \underline{q}_{S,c})(\overline{q}_{F,c} - \underline{q}_{F,c})} \int_{\underline{q}_{S,c}}^{\overline{q}_{S,c}} \int_{\underline{q}_{F,c}}^{\overline{q}_{F,c}} \Lambda(V_d(u, v))\, dv\, du, \tag{13}$$

where V_d is the choice index of eq. (5). In conviction units, the group's convictions are truncated normals, so eq. (13) has no closed form. I compute it by Gauss–Legendre quadrature with 16 nodes for each rank interval; with 64 nodes, the criterion changes by less than 0.01 and each margin by less than 0.01 points. The choice curves of Figure 3, eq. (7), average the same logit over the competing value's rank and over the card mix in the experiment; I compute them, and the margins, with 96 nodes.

For example, a respondent at the median of both convictions, where both indices and their product are zero, chooses the sustainable fund with probability $\Lambda(\delta_d)$: 64 percent on the control card and 80 percent on the rating cards.

A.2 Estimation and uncertainty

I estimate the model by choosing the parameters that make its predicted reporting shares and fund-choice rates match the observed data as closely as possible. This requires calculating the choice probability of each reporting group, eq. (13), and comparing those predictions with the observed frequencies.

The model is estimated using observed sustainable-fund choice rates for 60 respondent groups. Of these, 36 combine question order, pooled card category (control versus

any ESG-rating card), and the reporting category (open and closed combinations) for each topic: neither report, closed-ended endorsement only, or an open-ended mention. The remaining 24 groups distinguish late, middle, and early mentioners separately for each topic, question order, and pooled card category. These earliness groups overlap the reporting groups and provide additional information about choice differences among writers.

The model represents each group c by a conviction-rank interval for sustainability and another for fees:

$$B_{j,c} = [\underline{q}_{j,c}, \overline{q}_{j,c}], \qquad j \in \{S, F\}.$$

Here, $\underline{q}_{j,c}$ and $\overline{q}_{j,c}$ are the interval's lower and upper boundaries. The combination $B_{S,c} \times B_{F,c}$ describes the group's assigned sustainability and fee conviction ranges, and the group's predicted choice rate is eq. (13) on the group's card.

The eight choice parameters are

$$\theta_C = (\delta_0, \delta_1, \kappa_S^0, \kappa_S^1, \kappa_F, g_S, g_F, \kappa_{SF}),$$

with each skew restricted to $|g_j| < 3$. The reporting parameters, collected in θ_R, add four parameters per topic: the closed-ended and open-ended thresholds when each question is asked first, the closed-ended threshold after an open question, and the priming parameter η_j determining the open threshold after a closed-ended question. The priming rule places this threshold between the two first-position thresholds:

$$q_{jP} = q_{jC}^{1-\eta_j} q_{jO}^{\eta_j}, \qquad 0 \le \eta_j \le 1. \tag{14}$$

With $\eta_j = 0$, seeing the closed-ended list makes writing as easy as selecting a multiple-choice option; with $\eta_j = 1$, seeing the list has no effect on writing. Earlier questions can change later answers by bringing related considerations to mind (Tourangeau et al. 2000), and a list in particular brings up considerations that respondents would not have named spontaneously (Schuman et al. 1986). The parameter η_j is estimated separately for each topic, so the restriction $q_{jC} \le q_{jP} \le q_{jO}$ fixes only the order of the three thresholds and leaves the size of the priming effect free for each topic. This effect can differ between topics because a list helps most for values that respondents hold but rarely name unprompted (Laurent et al. 1995, Schuman et al. 1986); Appendix B.8 finds a large effect for fees and none for sustainability. All 16 parameters are estimated jointly.

The estimator uses 60 observed choice rates and eight reporting shares. Of the choice rates, 36 come from combinations of question order, pooled card group, and the three reporting intervals for each topic ($2 \times 2 \times 3 \times 3 = 36$). The other 24 distinguish

three mention-earliness groups for each topic, question order, and pooled card group ($3 \times 2 \times 2 \times 2 = 24$). The two pooled card groups are the control card and the three ESG-rating cards combined. The earliness groups add information about differences among respondents who mention a topic; their construction is described in Section B.2. The eight reporting shares distinguish two topics, two question formats, and two question orders.

Each part of the data has its own role. The eight reporting shares locate the thresholds; the fitted shares are within one percentage point of the observed ones. The thresholds fix where each reporting group lies on the conviction scale, so the differences between the choice rates of the groups, including early and late writers, determine the weights and the skews. The randomized rating cards, assigned after the value questions, add a second set of choice rates for the same groups and determine the intercept and the sustainability weight on the rating cards. Refits on subsets of the 60 choice rates show which data carry the weights and the skews. Without the writer groups, the fee weight is 0.46 instead of 0.40, and the fee skew is 0.02 instead of 0.33: the earliness groups carry the fee skew. Without the recognition-only groups, the fee weight falls to 0.28 and the fee skew rises to 0.84, and the standard error of the logarithm of the fee weight doubles. With the control card only, the parameters of the rating cards are not identified, the standard error of the logarithm of the fee weight rises by 50 percent, and the sustainability skew is 0.01: the rating cards carry the sustainability skew. In every refit, the sustainability weight stays between 0.64 and 0.73 on the control card and, where it is identified, between 0.86 and 0.92 on the rating cards, and the larger margin remains the same for each value.

Let $\widehat{p}_c$ denote an observed choice rate and $\widehat{r}$ the vector of observed reporting shares. Their model predictions are $p_c(\theta_C, \theta_R)$ and $r(\theta_R)$. I choose the parameters to minimize

$$Q(\theta_C, \theta_R) = \sum_{c=1}^{60} \left\{ \frac{\widehat{p}_c - p_c(\theta_C, \theta_R)}{\widehat{\mathrm{se}}(\widehat{p}_c)} \right\}^2 + \{\widehat{r} - r(\theta_R)\}' \widehat{\Omega}_r^{-1} \{\widehat{r} - r(\theta_R)\}.$$

The first term measures errors in predicted choice rates, giving greater weight to rates estimated more precisely. The second measures errors in reporting shares. The covariance matrix $\widehat{\Omega}_r$ accounts for correlations between reporting shares calculated from the same respondents.

Some respondents contribute to both the 36 reporting-interval choice rates and the 24 earliness choice rates. These statistics therefore have correlated estimation errors, which the diagonal choice weights do not fully account for. Consequently, the 68 statistics and 16 parameters give only a nominal count of 52 residual degrees of freedom; Q is treated as a descriptive fit measure. Alternative weighting methods give similar results:

the fee weight ranges from 0.40 to 0.43, the sustainability skew from -0.76 to -0.57, and each margin moves by at most 0.7 points.

To calculate uncertainty, I draw 2,000 bootstrap samples by resampling respondents with replacement within each combination of question order and the four original randomized cards. In every sample, I reconstruct reporting shares, earliness groups, choice rates, and weights, then re-estimate all parameters, including the reporting thresholds. The 95% percentile-bootstrap intervals use the 2.5th and 97.5th percentiles of the resulting estimates. These intervals can be less reliable when estimates reach imposed parameter limits. This happens in 12 percent of the samples, in most of them for the fee skew: a small resampled cell has a choice share of exactly zero or one, gets the floor of the cell standard error and dominates the criterion, and the skews bend the choice curve toward it, which widens the tails of the weights and the skews. Computing the cell standard errors from the smoothed share $(x + 0.5)/(n + 1)$ of a cell with x sustainable choices among n respondents removes most of these samples (1 percent remain) and the tails (Table A.1).

Table A.1: Choice and priming estimates

Parameter	Estimate	Interpretation
κ_F	0.404	Fee weight
κ_S^0	0.692	Sustainability weight, control card
κ_S^1	0.907	Sustainability weight, rating cards
g_S	-0.764	Sustainability skew
g_F	0.328	Fee skew
κ_{SF}	0.046	Product coefficient
δ_0	0.574	Intercept, control card
δ_1	1.389	Intercept, rating cards
η_S	1.000	Sustainability priming position
η_F	0.793	Fee priming position
Q	38.823	68 moments; 52 nominal residual degrees of freedom

Note: This table shows the model estimates. The weights measure how much one standard deviation of conviction moves the log odds of choosing the fund its value favors near the median, where each index has slope one; the skews bend the indices, eq. (4); the product coefficient multiplies the product of the two indices; the intercepts are the log odds of choosing the sustainable fund at median convictions. The 95% bootstrap intervals are $[0.04, 1.03]$ for κ_F, $[0.38, 1.30]$ for κ_S^0, $[0.56, 1.23]$ for κ_S^1, $[-2.79, -0.10]$ for g_S, $[-0.58, 3.00]$ for g_F, and $[-0.35, 0.20]$ for κ_{SF}, based on 2,000 respondent-level resamples that repeat the full estimation; the upper limit for g_F is its bound. With cell standard errors from the smoothed shares, they are $[0.19, 0.61]$, $[0.41, 1.06]$, $[0.61, 1.14]$, $[-1.89, -0.20]$, $[-0.32, 1.19]$ and $[-0.11, 0.20]$. The fit criterion Q is descriptive.

B Model Assumptions and Validation

Table B.1 lists the assumptions examined below and the subsection that examines each.

Table B.1: Model assumptions examined in this appendix

Assumption	Section
A. Survey answers and conviction	
Reporting follows conviction: writers hold the strongest convictions, and everyone who writes about a topic would also tick it.	B.1, B.4
Among writers, those who mention a topic earlier in their answer hold stronger convictions.	B.2
Seeing the list first can make writing about a topic easier, but not easier than ticking.	B.7, B.8
B. Conviction	
Sustainability and fee convictions are independent.	B.3
The choice index is a skewed normal score of conviction, with one skew per value.	B.4
C. Fund choice	
The fee weight is the same on every card.	B.4
The skews are the same on both card designs.	B.4
Conviction enters the choice directly, without a hurdle.	B.5
Writing or ticking a value does not itself commit respondents to it.	B.6.1, B.1.1
Stated values relate to fund choice in the same way in both question orders.	B.6.1, E.4.3
Reports predict fund choice in the same way under each of the three ESG-rating cards.	B.6.2

Note: Each subsection describes how the assumption is checked and what the check shows. The checks do not establish that the assumptions hold exactly. Appendix B.7 also compares the model with the OLS coefficients; Appendix A describes the group choice probabilities, estimation weights, and the bootstrap.

B.1 Reporting intervals and paired responses

The model constructs three conviction ranges from the overall shares of open and closed endorsements. With ordered reporting thresholds, anyone who mentions a topic spontaneously would also select it in the closed-ended question. Actual responses sometimes depart from this pattern: Table B.2 shows respondents who mention a topic but do not select it, labelled W. The estimator retains these respondents and groups them with everyone else who mentions the topic. Its upper writer interval therefore includes both R and W from the table.

Consequently, the model's intervals do not need to match the sizes of the observed response groups. For example, in the closed-first arm, 41.0 percent select sustainability and 33.7 percent mention it in the subsequent open-ended response. A nested partition matching these totals would assign 7.3 percent (i.e., the difference between open and closed-ended) to the recognition-only interval, although the actual recognition-only group contains 13.8 percent of respondents. The model uses the observed choice rate of

recognition-only respondents (i.e., the 13.8 percent in this example) to estimate behavior in its recognition-only conviction interval, even though that interval can differ in size from the observed group.

The alternative model in Appendix B.1.1 allows open-only responses (i.e., write about a topic but not select it in the closed-ended question). Its estimates preserve the finding that switching from open to closed-ended questions has the larger effect for fees (i.e., more weak values), while negative screening of the competing value has the larger effect for sustainability (by screening out low fee importance). Because the alternative keeps the choice equation of the main model, it shows how much the results change from allowing respondents to mention a topic without selecting it in the closed-ended question: fee broadening is 1.6 points smaller, and each other effect changes by less than one percentage point.

Table B.2: Paired open and closed-ended response patterns

Order arm	Topic	U	K	R	W (open only)
Open first	Fees	45.2	38.3	15.1	1.4
Open first	Sustainability	45.9	18.7	32.7	2.7
Closed-ended first	Fees	42.7	31.5	22.2	3.6
Closed-ended first	Sustainability	52.5	13.8	27.2	6.5

Note: Entries are percentages within arm and topic. U means neither report; K means closed-ended only; R means both; W means open only. The main estimator matches marginal shares and treats all writers as the upper aggregate interval.

B.1.1 Alternative model with individual stochastic reporting

Section 6.3 refers to an alternative model with individual stochastic reporting that permits nonnested reports. This subsection records that model, its likelihood fit, and additional variants. The model retains

$$Z_j \sim N(0,1), \qquad q_j = \Phi(Z_j),$$

and the choice step of Section 6.2, but replaces deterministic reporting thresholds with ordered logistic gates:

$$\Pr(T_{jm} = 1 \mid q_j, \omega) = \Lambda[\iota_{jm} + \nu_j \operatorname{logit}(q_j) + \upsilon_{jm}(\omega)]\,,$$
$$\iota_{jC} = \iota_{jO} + \psi_j, \qquad \psi_j > 0,$$

where T_{jm} indicates a report of topic j in format m and $\upsilon_{jm}(\omega)$ shifts each gate by question order. Thus, apart from the order shifts, closed-ended reporting is easier at every rank, while realized reports need not be nested. The slope $\nu_j > 0$ puts the two gates of a topic on one logit-rank scale. A single slope for both topics lowers twice the

log likelihood by 43.5 for one parameter fewer. The model with one slope per topic is retained.

Choice depends on the two convictions and the randomized card through eq. (6). The pre-choice variant uses the four pre-choice reports and choice, and it excludes direct report terms from utility so that reports remain measurements of conviction rather than additional preferences. Against the saturated likelihood, its deviance is 145.10 on 108 nominal residual degrees of freedom ($p = 0.010$), so the model is not a complete fit to every response pattern.

In the pre-choice model, the fee weight is $\widehat{\kappa}_F = 0.52$, the sustainability weights are 1.4 and 2.2 times as large on the control and the rating cards, and the skews are -0.17 for sustainability and 0.13 for fees. Appendix B.4 shows that the main model's sustainability skew also shrinks when the respondents who write without ticking count as tick-only. Because the reports are no longer a function of conviction, this model can add direct report terms to the choice, but only through the form of its gates, which treat a written but unticked answer like a ticked but unwritten one. For both topics, these terms have the wrong sign for a commitment effect; Appendix B.6.1 gives a design-based check.

Finally, the pre-choice stochastic model moves the paper's economic effects in the same directions. Target broadening changes the fee and sustainability gaps by +7.8 and +1.2 points, against +9.4 and +0.7 in the main model. Competing-value screening changes the two gaps by -2.2 and -2.6 points, against -2.6 and -2.4. With the gates on the conviction itself, $\Lambda(\iota_{jm} + \nu_j Z_j + \upsilon_{jm})$, as for the ordered indicators of hybrid choice models (Daly et al. 2012), the four changes are +7.7, +1.2, -2.2 and -2.6 points.

B.2 Earliness and conviction ranks

The model uses how early a topic first appears in a response compared to other mentioned topics to rank topic importance within a respondent. For each topic and question order, I rank respondents who mention the topic by how early the first topic-related word appears relative to their answer's length. I divide this ranking into three groups, chosen to be as equal in size as possible without splitting identical earliness scores. "Late," "middle," and "early" refer to these groups in the earliness ranking.

For example, the 258 respondents who mention fees when asked the open-ended question first would form groups of 86 if divided into exact thirds. Keeping identical scores together instead produces 88 late, 80 middle, and 90 early mentioners. The model's estimated mention threshold places all three groups above the 84th percentile of fee conviction. The estimator then divides the range from 84 to 100 according to their observed group sizes, assuming that earlier-mentioning groups have higher conviction

than later mentioners. Late mentioners account for 34% of these writers and therefore receive the lowest 34% of this range, approximately percentiles 84–89. Middle mentioners receive percentiles 89–94, and early mentioners receive percentiles 94–100.

Comparing the earliest-mentioning group with the other two groups and adjusting for the card shown, earlier mentions are positively associated with value-consistent choices. The evidence is strong for fees ($p = 0.0003$) and weaker but still significant at the 10% level for sustainability ($p = 0.065$). Appendix Table B.3 reports the corresponding choice shares and trend tests separately by card condition. These findings are consistent with earliness signaling conviction, particularly for fees, but do not validate the exact rank intervals imposed by the model.

I re-estimate the model without using how early a topic appears in the written answer. All respondents remain in the analysis. The model still uses their topic endorsements and fund choices, but drops the choice rates calculated separately for the earliness groups. There are 24 such rates: three earliness groups, two topics, two question orders, and two pooled card groups ($3 \times 2 \times 2 \times 2 = 24$). The latter distinguish the control card from the three ESG-rating cards, which are combined for estimation.

The main findings remain (Table B.4). For fees, broadening endorsement increases the value–action gap more than screening for competing sustainability concerns reduces it. For sustainability, screening for competing fee concerns reduces the gap more than broadening endorsement increases it. In this refit, the 95 percent bootstrap interval of the difference between the larger and the smaller effect excludes zero for sustainability; for fees, its lower end is close to zero.

The fee weight barely depends on the use of earliness: without it, the estimate is 0.46, with a 95 percent interval of $[0.28, 0.82]$. The fee skew does: without earliness, it is 0.02 instead of 0.33 (Appendix A.2).

Table B.3: Raw first-token-earliness gradients

Topic	Card	T1	T2	T3	Trend p
Sustainability	Control	65.4 (81)	75.3 (81)	76.9 (78)	0.103
Sustainability	Rating	85.3 (272)	84.7 (268)	88.8 (278)	0.223
Fees	Control	61.8 (34)	58.1 (43)	80.0 (40)	0.084
Fees	Rating	42.5 (120)	46.7 (107)	61.2 (116)	0.004

Note: Entries are raw value-consistent choice percentages with cell sizes in parentheses. T1–T3 are tie-respecting earliness groups, pooled across order arms only for display. The estimator uses 24 arm-specific cells and their empirical widths. Fee-writer cells exclude respondents who also write about sustainability, keeping the two writer blocks disjoint. No monotonicity is imposed on the displayed means.

Table B.4: Identification sensitivities

Specification	Moments	Q	κ_F	κ_S^0	κ_S^1	Skews F/S	Target F/S	Screen F/S
Main model	68	38.82	0.40	0.69	0.91	0.33/−0.76	9.4/0.7	−2.6/−2.4
No earliness moments	44	20.74	0.46	0.64	0.92	0.02/−0.85	7.7/0.6	−2.7/−3.0

Note: This table shows how the estimates change when the earliness information is removed. Paired entries report fees/sustainability. Target and screening effects are percentage points. In the main model, the 95% bootstrap intervals are $[5.1, 21.4]$ and $[0.1, 1.3]$ for the target effects, $[-3.8, -1.7]$ and $[-5.0, -0.2]$ for the screening effects, and $[2.5, 18.8]$ and $[-4.7, 0.7]$ for the net changes, +6.8 and −1.7 points; with cell standard errors from the smoothed shares, the intervals of the net changes are $[2.7, 11.8]$ and $[-3.8, 0.1]$. Without the earliness moments, the net changes are +5.0 and −2.3 points, with intervals of $[0.1, 10.7]$ and $[-4.9, -0.4]$. The no-earliness specification uses fewer moments, so its fit criterion Q is not directly comparable.

B.3 Independence of the two convictions

The main model assumes that sustainability and fee conviction ranks are independent. Knowing that someone has strong sustainability conviction therefore provides no information about their fee conviction. I examine this assumption using the reported answers and by allowing dependence in the structural model.

The correlations between sustainability and fee reports are close to zero for closed-first answers ($\phi = 0.027$, $p = 0.316$), closed answers following an open question ($\phi = -0.006$, $p = 0.839$), and open answers following a closed question ($\phi = -0.035$, $p = 0.189$). Unprimed open answers show a small negative association ($\phi = -0.098$, odds ratio 0.55), significant at the 1 percent level after Holm adjustment for multiple testing. A test using all four paired response patterns also rejects independence in the open-first arm, although the association is small (Cramér's $V = 0.065$). This could reflect respondents concentrating their written answer on one topic, even when they care about both. Dependence between reports therefore need not imply the same dependence between underlying convictions.

To allow related convictions, I use a Gaussian copula, as for the survey in Appendix E.3. This is a one-parameter formula linking the two rank distributions. Its joint density is

$$c_\rho(q_S, q_F) = \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{\rho^2 \left(x_S^2 + x_F^2\right) - 2\rho\, x_S x_F}{2\left(1-\rho^2\right)} \right), \tag{15}$$

where $x_j = \Phi^{-1}(q_j)$, Φ is the standard normal distribution function, and $\rho \in (-1, 1)$. Each rank remains uniformly distributed. Zero means independence; positive values associate higher sustainability ranks with higher fee ranks, while negative values associate higher sustainability ranks with lower fee ranks. The parameter ρ is the correlation between the normal scores x_S and x_F, which in this model are the convictions themselves, so the copula can represent dependence of any strength.

The choice equation retains its form, eq. (6). For a choice cell c, let $\mathcal{B}_c$ denote its sustainability and fee rank intervals. Its predicted choice rate becomes

$$P_c(\vartheta, \rho) = \frac{\iint_{\mathcal{B}_c} \Lambda(V_{d_c}(q_S, q_F)) c_\rho(q_S, q_F)\, dq_S\, dq_F}{\iint_{\mathcal{B}_c} c_\rho(q_S, q_F)\, dq_S\, dq_F}, \tag{16}$$

where ϑ collects the remaining structural parameters. This expression averages choice probabilities over the cell, allowing some combinations of convictions to occur more frequently than others. The integrals are calculated by numerical quadrature, without simulation.

For each ρ, I re-estimate all remaining parameters, including the reporting thresholds, using the same 68 moments and weights:

$$Q_{\text{prof}}(\rho) = \min_{\vartheta} Q(\vartheta, \rho), \qquad \widehat{\rho} = \arg \min_{\rho \in (-1,1)} Q_{\text{prof}}(\rho).$$

The reports give a second estimate of ρ that does not depend on the choice equation. In the model, each report is a threshold on a conviction rank, so how often each combination of the two topics' unengaged, recognition-only, and written intervals occurs depends only on the thresholds and on ρ. Estimating ρ by maximum likelihood from these combinations, with a common ρ in both arms and thresholds at the observed reporting shares (the polychoric correlation of the intervals), gives $\rho = -0.07$, with a 95% likelihood-ratio interval of $[-0.12, -0.02]$. Adding this likelihood-ratio statistic to $Q_{\text{prof}}(\rho)$ and minimizing over ρ gives the same value.

Table B.5 reports both alternatives. With ρ fixed at the value implied by the reports, the fee weight changes from 0.40 to 0.39, and each margin changes by less than 0.2 percentage points. The choices alone pin down ρ only loosely: between $\rho = -0.3$ and $\rho = 0.5$, Q_{prof} varies by at most 2.4. Its minimum is at $\widehat{\rho} = -0.03$, at the upper end of the interval implied by the reports (the likelihood-ratio statistic is 3.4), and the fit criterion falls by less than 0.01. The choices and the reports thus agree on weak dependence, which barely changes the effects, so the main model keeps independent ranks.

Table B.5: Sensitivity to cross-topic rank dependence

Rank specification	ρ	Q	κ_F	κ_S^0	κ_S^1	Target F/S	Screen F/S
Independent ranks	0.00	38.83	0.404	0.692	0.908	9.38/0.68	−2.62/−2.41
Implied by the reports	−0.07	38.85	0.389	0.678	0.892	9.28/0.67	−2.48/−2.39
Best fit to the 68 moments	−0.03	38.82	0.398	0.687	0.902	9.35/0.68	−2.57/−2.40

Note: This table compares the main model with independent conviction ranks and two models in which a Gaussian copula links the ranks. In the second row, ρ is fixed at the polychoric correlation of the two topics' reporting intervals; in the third, it is estimated from the 68 moments. All rows re-estimate the other parameters on the same 68 moments; they integrate over the copula with 64 nodes for each rank, so the criterion of the first row differs from that of Table A.1 in the second decimal. Target and screening effects are in percentage points, reported as fees/sustainability.

B.4 The latent scale

The main model applies Tukey's g-transformation to the normal score $\Phi^{-1}(q)$ of each conviction rank, with one skew per value, and lets the two indices interact through their product (Section 6.1). With ordinal answers and a single choice, only the ranking of convictions and the shape of the choice along this ranking are identified, so the skews estimate this shape, not the distribution of conviction itself. Figure B.1 draws the two indices: the sustainability index is bounded above, so the strongest sustainability convictions add little to the choice, whereas the fee index has a long upper tail. Table B.6 reports alternatives, each re-estimated on the same 68 moments.

Without skews, the normal score fits clearly worse ($Q = 53.49$ with the product term and 53.56 without it, against 38.82). Because the criterion is descriptive, I calibrate the gain by simulation: 1,000 samples drawn from the model with normal scores and without the product term on the actual design, each refitted with the normal and the skewed scores, both without the product term.[14] A drop as large as that from the two skews occurs in 1.1 percent of the samples, so with smoothed cell standard errors the data reject the normal shape. The skew of sustainability is negative and differs from zero; the skew of fees is positive but does not (Table A.1). In each of the three other surveys, the skew of the more salient value is also negative and below that of the less salient value; the skews are jointly significant in the Zurich survey but not in the other two (Appendices E.3 and E.4).

The size of the sustainability skew depends on how the 140 respondents who write about sustainability without ticking it are counted. On the rating cards, they choose the sustainable fund less often than the other writers and than respondents who tick it

[14] Each sample keeps every respondent's question order and choice card and draws the conviction ranks, the reports and the fund choice. A simulated cell whose share is exactly zero or one would get the floor of the cell standard error and could dominate the criterion, so the simulations compute cell standard errors from the smoothed share $(x + 0.5)/(n + 1)$ of a cell with x sustainable choices among n respondents. With the floor, as in the estimation, such cells dominate the simulated drops, and a drop as large as the observed one occurs in 13.5 percent of the samples.

without writing (66 against 89 and 82 percent). Left out, they leave a skew of -0.54; counted as tick-only, as in the stochastic-reporting version of Appendix B.1.1, -0.30, which no longer differs from zero. The margins change by at most 0.4 points.

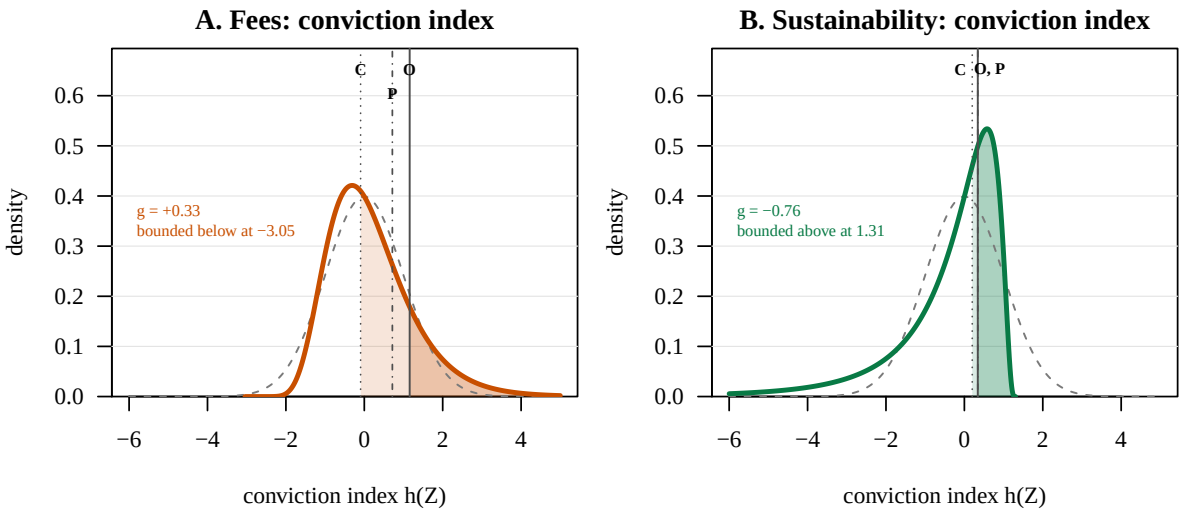
The logistic score $\text{logit}(q)$ fits about as well as the normal score ($Q = 54.04$). The log rank $\log q$, which compresses the strongest convictions, fits worse ($Q = 56.02$), and the Gumbel score $-\log(-\log q)$, which stretches them, fits clearly worse ($Q = 68.43$); all three keep the product term. The beta-logistic score $\text{logit}(q^{\gamma_j})$, with one shape parameter for each value instead of a skew, fits about as well as the main model ($Q = 38.53$). The product term adds little in this survey: without it, the criterion rises by 0.5, and each margin changes by less than 0.4 points. The model keeps it because it matters in the survey of Andre et al. (2022), where respondents who hold both channels predict a change in either direction less often (Appendix E.3), and the specification is the same in all four surveys.

The margins do not depend on the scale: target broadening remains the larger effect for fees and screening the larger effect for sustainability in every row, although the sizes change. The log rank gives the smallest fee broadening, 3.3 points; the normal score raises the sustainability broadening from 0.7 to 1.3 points.

Two restrictions on the weights test the card design. Estimated separately by card, the fee weight does not differ: the rating-card weight is 0.84 times the control-card weight, and the criterion improves by 0.7. A common sustainability weight for both card designs raises the criterion by 3.3 for one parameter fewer. A single index, with equal weights for the two convictions on the control card, raises it by 4.9.

The skews can also differ between the card designs, because the rating cards show the funds' sustainability rating. Estimated separately by card, the sustainability skew is -0.04 on the control card and -1.04 on the rating cards; the difference has a standard error of 0.54, and the criterion falls by 3.4 for one parameter more. The fee skew does not differ between the cards (0.41 and 0.30). The flattening of the sustainability index is thus concentrated on the rating cards: there, respondents who only tick sustainability choose the sustainable fund almost as often as writers (82 against 86 percent), whereas on the control card they lie closer to non-reporters (56 against 45 and 73 percent). The rating itself may produce this flattening, because it gives respondents with a mild concern for sustainability a reason to choose the sustainable fund; the model does not separate this from salience. The margins change by less than 0.2 points. As in the other surveys, the main model keeps one skew per value.

Figure B.1: The conviction index of each value



Target-reporting thresholds: C = closed-ended; P = primed open; O = unprimed open. Dashed: no skew (standard normal).

Note: The panels show the distribution of the conviction index $h_j(Z_j)$ of eq. (4) at the estimated skews; the dashed line is the index without skew, the standard normal. Shading marks respondents who tick the value without writing about it (light) and respondents who write about it (dark); vertical lines denote the closed-ended (C), primed open (P), and unprimed open (O) reporting thresholds. On the log-odds scale of Figure 3, a choice curve is close to a straight line in its value's index, so a bounded index flattens the curve among strong convictions and a long upper tail steepens it.

Table B.6: Alternative latent scales and choice steps on common moments

Specification	Par.	Q	Target F/S	Screen F/S
A. Latent scale				
Skewed normal score and product (main model)	16	38.82	9.4/0.7	−2.6/−2.4
Skewed normal score, no product	15	39.33	9.1/0.7	−2.5/−2.7
Normal score and product	14	53.49	7.6/1.3	−2.1/−3.1
Normal score, no product	13	53.56	7.7/1.3	−2.1/−3.0
Logistic score and product	14	54.04	8.1/1.3	−2.0/−3.0
Log rank and product	14	56.02	3.3/0.6	−2.7/−4.1
Gumbel score and product	14	68.43	9.5/1.7	−1.7/−2.5
Beta-logistic score and product	16	38.53	9.0/0.7	−2.6/−2.6
B. Weights				
Fee weight by card	17	38.13	9.4/0.7	−2.6/−2.4
Common sustainability weight	15	42.14	9.4/0.6	−2.7/−2.4
Single index	15	43.67	9.0/0.4	−2.7/−3.2
C. Choice step				
Hurdle version	13	61.05	6.2/1.4	−2.0/−4.6

Note: All specifications fit the same 68 moments using the same weights and reporting restrictions. Panel A replaces the main model's skewed normal score of the conviction rank in the choice index; the skewed scores apply Tukey's g-transformation to it. Panel B restricts or frees the main model's weights by card; the single index sets the control-card sustainability weight equal to the fee weight. Panel C is the hurdle version of Appendix B.5. Paired entries are fees/sustainability. Target and screening effects are percentage points.

B.5 The hurdle version

In the main model, conviction enters the choice directly: the stronger a conviction, the more its value weighs. In a hurdle version, conviction instead decides whether a value enters the decision at all, and the values that enter decide the choice, as in a hurdle model (Cragg 1971) or a model of endogenous attribute attendance (Hole 2011). Each value j enters with probability $q_j^{\gamma_j}$, which rises with the conviction rank q_j. When one value enters alone, the investor chooses the fund it favors; when both enter, the investor trades off the funds' attributes as the card shows them; when neither enters, the investor chooses the conventional fund on the control card and the sustainable fund with an estimated probability on the rating cards. The hurdle version uses the same 68 moments, weights and reporting thresholds as the main model, and has 13 parameters, three fewer.

It fits worse: its criterion is 61.05, against 38.82, and against 53.56 for the model with normal scores and without the product term, which has as many parameters. Because the criterion is descriptive, I calibrate the difference to the latter by simulation, as in Appendix B.4: 1,000 samples drawn from each of the two specifications on the actual design, each refitted with both. A difference as large as the observed one occurs in 1 percent of the samples simulated from the hurdle version but is typical of samples simulated from the model with normal scores (68 percent).[15] The split of the format effects does not depend on this choice: the hurdle version gives +6.2 and −2.0 points for fees and +1.4 and −4.6 points for sustainability (Table B.6). As Vij and Walker (2016) argue for latent variables in choice models, the case for a hurdle would be what it separates, whether a value enters the choice, not its fit. With one choice per respondent, however, the data can hardly separate a value that stays out of the choice from one that enters with a low weight (for repeated choices, see Hess et al. 2013).

The order in which respondents mention the two values gives a check that the estimator does not use. Among the 183 respondents who write about both, those who mention sustainability first choose the sustainable fund more often than those who mention fees first (80 against 68 percent). The main model predicts 78 and 68 percent, through each value's own earliness: the estimator places writers who mention a value earlier at stronger convictions. The hurdle version predicts 75 and 68 percent.

B.6 Pooling across question orders and rating cards

The preferred model uses the same choice rules in both question orders and pools the three ESG-rating cards. Two reduced-form checks examine these restrictions.

[15] As in Appendix B.4, the simulations smooth the cell standard errors. Without the smoothing, a difference as large as the observed one occurs in 20 percent of the samples simulated from the hurdle version and in 68 percent of those simulated from the model with normal scores.

B.6.1 Question order

I examine whether question order changes the relationship between reported values and subsequent choice. These checks use all 3,059 respondents.

Let $Y_i = 1$ indicate sustainable-fund choice and $O_i = 1$ indicate that the open question was asked first. Let $T_i \in \{1, 2, 3, 4\}$ identify the original randomized card, with $T_i = 1$ denoting the control card. These regressions distinguish all four cards. Define $X_{ijm} = 1$ if respondent i endorses topic $j \in \{S, F\}$ in format $m \in \{O, C\}$.

I first compare overall sustainable-fund choice across question orders. The shares are 65.6 percent in the open-first arm and 67.8 percent in the closed-first arm. A two-sided test of equal proportions, without continuity correction, does not reject equality ($\chi^2(1) = 1.734$, $p = 0.188$).

Asking about an intention can change the behavior that follows (Morwitz et al. 1993, Sprott et al. 2006). Because the question order shifts the reports, this comparison also bounds such a commitment effect of answering, in which writing or ticking a value would move the choice toward the fund the value favors. Seeing the list first raises fee writing by 9.3 points, and writing first raises sustainability ticking by 10.3 points; the other two reports change by less than 2 points. If each report moved the choice by c percentage points, the closed-first arm would choose the sustainable fund $0.22c$ points less often than the open-first arm; it chooses it 2.2 points more often. Assuming that the order acts on the choice only through the reports, the estimate of c is negative, and the upper end of its 95 percent bootstrap interval is 3.7 points. This bound applies to a common effect of every report. For fee writing alone, the upper end is 9.4 points; for sustainability writing alone, the design gives no bound, because the order hardly changes sustainability writing. Such an effect of writing would raise the broadening margins, which compare writers with respondents who only tick.

I then estimate the logistic regression

$$\begin{aligned}\operatorname{logit}\Pr(Y_i = 1 \mid O_i, T_i, \mathbf{X}_i) = \alpha_{T_i} + \delta O_i \\ + \sum_{j\in\{S,F\}} \sum_{m\in\{O,C\}} (\beta_{jm} + \theta_{jm} O_i) X_{ijm},\end{aligned} \tag{17}$$

where $\mathbf{X}_i$ contains the four report indicators and α_{T_i} allows a separate intercept for each card. The coefficients θ_{jm} measure whether the association between each report and choice differs across question orders. I test

$$H_0 : \quad \theta_{SO} = \theta_{FO} = \theta_{SC} = \theta_{FC} = 0.$$

A likelihood-ratio test compares this specification with one that retains the card intercepts and the order indicator but omits all four order-by-report interactions. Allowing the report coefficients to differ across orders does not significantly improve the regression ($\chi^2(4) = 0.897$, $p = 0.925$). This test is weak: the model itself implies small differences between the orders, because the list primes additional fee writers in the closed-first arm. Appendix E.4.3 gives a model-based check, predicting each order's gaps from the other.

B.6.2 Pooling of rating cards

Using the variable definitions in Section B.6.1, I examine whether the three ESG-rating cards can be pooled. I estimate a separate linear probability model for each reporting format:

$$Y_i = a_{m,T_i} + d_m O_i + b_{Sm,T_i} X_{iSm} + b_{Fm,T_i} X_{iFm} + \varepsilon_{im}, \qquad m \in \{O, C\}. \tag{18}$$

Both the intercept and the two endorsement coefficients may differ across all four cards. For each topic and format, I test whether the endorsement coefficient is equal across the three rating cards:

$$H_0 : \quad b_{jm,2} = b_{jm,3} = b_{jm,4}.$$

These are four separate Wald tests, each imposing two restrictions and using HC2 heteroskedasticity-robust standard errors. For closed reports, the p-values are 0.711 for sustainability and 0.756 for fees; for open reports, they are 0.783 and 0.176, respectively.

These tests find no statistically significant differences in the endorsement coefficients across the three ESG-rating cards. They concern how endorsements predict choice. Because the card intercepts remain unrestricted, they do not establish that the three rating cards generate identical overall choice rates. They also do not directly test equality of the model's weights across the three rating cards.

B.7 Reporting fit and OLS comparison

This subsection examines how well the fitted model reproduces observed reporting shares and the relationship between endorsements and fund choice. Reporting shares enter estimation directly, while the OLS coefficients provide a comparison calculated after estimation using the same data.

B.7.1 Reporting shares

Table B.7: Observed and jointly fitted reporting shares

Topic	Format and order	Observed (%)	Fitted (%)
Sustainability	Closed-ended first	41.03	41.38
	Open after closed-ended	33.73	34.59
	Open first	35.40	34.59
	Closed-ended after open	51.37	51.17
Fees	Closed-ended first	53.68	53.81
	Open after closed-ended	25.77	26.01
	Open first	16.49	16.33
	Closed-ended after open	53.42	53.24

Note: The shares are fitted jointly with the 60 choice moments. For each topic, the primed-open threshold obeys eq. (14); the closed-ended-after-open threshold is free. Sustainability priming reaches the no-movement boundary ($\widehat{\eta}_S = 1.000$), so the two fitted open shares coincide. The reporting component of the minimum-distance criterion is 1.176.

B.7.2 Connection to the OLS estimates

The model reproduces the broad OLS pattern, but shows a smaller effect on how much alignment fee endorsements lose when moving from open to closed questions. Table B.8 compares the observed OLS coefficients with those implied by the fitted structural model. The OLS coefficients are not directly used as estimation targets. Their model counterparts are calculated afterward from the fitted choice probabilities, without simulation. This comparison checks how well the model reproduces the regression associations in the estimation sample. All coefficients discussed below are expressed in percentage points.

For fees, the observed coefficient falls from 17.7 under unprimed open reporting to 12.0 under closed-first reporting. The corresponding model-implied coefficients are 16.7 and 11.8. Both show a weaker association between fee endorsement and low-fee choice under closed reporting, but the decline is 5.7 points in the OLS regressions and 4.9 points in the structural model. For primed-open reports, the observed coefficient is 12.1, compared with 14.5 in the model. The observed differences between the fee coefficients across reporting conditions are imprecisely estimated. For sustainability, the coefficients are comparatively stable: the observed values are 25.7, 23.0, and 27.8 for unprimed open, primed open, and closed-first reports, compared with model-implied values of 25.7, 25.7, and 27.3.

Across all eight binary-report coefficients, the root-mean-square discrepancy between model and data is 1.6 percentage points. Table B.8 also compares the recognition-only and recalled-group coefficients from columns (5)–(6) of Table 4. The largest discrepancy concerns the difference in sustainable-fund choice between recalled and unengaged respondents in the closed-first arm: 28.1 points in the model and 31.6 points in the data. This observed contrast also has substantial estimation uncertainty.

Table B.8: Observed and model-implied OLS coefficients

Topic	Coefficient	Closed-ended first		Closed-ended after open	
		Observed	Model	Observed	Model
Fees	Closed-ended report	0.120	0.118	0.134	0.118
	Recognition-only vs. U	0.095	0.069	0.089	0.081
	Recalled vs. U	0.169	0.171	0.234	0.203
Sustainability	Closed-ended report	0.278	0.273	0.316	0.301
	Recognition-only vs. U	0.239	0.231	0.253	0.256
	Recalled vs. U	0.316	0.281	0.349	0.322

Note: Value-consistent orientation: the fee entries are the negative of the high-fee-choice coefficients in Table 4. U is the omitted unengaged group; the observed engagement regressions include the write-only controls. The model is estimated on the 60 choice cells and eight reporting shares, not these OLS slopes. The remaining binary-report comparisons are observed/model 0.177/0.167 for unprimed-open fees, 0.121/0.145 for primed-open fees, 0.257/0.257 for unprimed-open sustainability, and 0.230/0.257 for primed-open sustainability.

B.8 Comparison with the raw data

This subsection compares the structural results of Section 7.2 with the raw data. The measured path of eq. (12) needs respondents who write about sustainability, or not, and tick fees, or not. Both question orders contain them, and neither answer depends on its position: 35 and 34 percent of respondents write about sustainability, and 54 and 53 percent tick fees, when the question comes first and second. Combining the two orders, broadening raises the fee gap by 12.5 points and the screen lowers it by 3.7, against 9.4 and 2.6 in the model; for sustainability, the screen lowers the gap by 3.6 points and broadening raises it by 0.7, against 2.4 and 0.7 (Appendix Table E.9). Broadening dominates for fees in 98 percent of the bootstrap samples and screening for sustainability in 87 percent, against 99 and about 90 percent in the model. Each order alone gives the same mixed gap for sustainability (12.3 and 11.9 percent) and, within sampling error, for fees (50.0 and 54.4 percent). The other path would need sustainability ticks and fee mentions, which both depend on position. Counted as in the model, the measured margins add up to the changes of 8.9 and -2.9 points reported below.

The paired answers, which change the format of one value at a time within a question-order arm (Section 5), give a split for sustainability within sampling error of the model's: $+2.4$ and -3.5 points in the open-first group, counted as in the model, against the model's $+1.8$ and -2.5 at that group's reporting thresholds (Appendix Table E.9). These comparisons use second-position answers, which the preceding question may influence. For fees, the paired answers give a larger offset from the competing value than the model's first-position split (-8.3 points in the open-first group, against -2.6). A likely reason is question position: after writing, 51% of respondents tick sustainability, compared with 41% when the closed-ended question comes first, so the second-position screen excludes more respondents. At the reporting thresholds of the open-first group, the

model gives an offset of −7.0 points. Fee ticking does not depend on position (53–54%), which is consistent with the closer agreement for sustainability.

The structural decomposition compares reporting rules under a common choice model. For fees, its net gap increase is 6.8 percentage points, compared with 10.8 points in the raw between-arm comparison (Section 5.1), which also changes question order; for sustainability, its decrease is 1.7 points, compared with 5.9. The model thus reproduces the direction of both raw changes with smaller estimates, but the differences are within sampling error: bootstrapping the raw and model changes jointly over the same 2,000 resamples, the 95% interval of their difference is $[-8.5, 10.6]$ points for fees and $[-9.1, 1.2]$ for sustainability. The same holds for the one-value gaps of Table 3: the model's changes, +8.5 points for fees and +0.8 for sustainability, are within sampling error of the raw changes, +14.5 and −1.0 (Appendix C.1).

The raw and model gaps also differ in how they treat respondents who write about a topic without selecting it from the list. The model is estimated with these respondents among the writers, and it assumes that every writer would also select the topic (Section 6.3); its closed-ended gaps therefore count them as reporters of the topic, whereas the raw closed-ended gaps count them as non-reporters.[16] When the raw data count them as closed-ended reporters, the changes are 8.9 points for fees and −2.9 for sustainability, and the model's changes have the same directions. The differences between the raw and model changes, 2.1 and −1.2 points, have 95% intervals of $[-10.6, 8.7]$ and $[-6.4, 4.0]$. Most of the fee difference lies at the open endpoint: among respondents who write about fees but not about sustainability, the model predicts a gap of 41.8%, whereas the observed gap is 39.7%.

As an out-of-sample check, a simpler version of the model, fitted only to the paired answers of the open-first group, predicts the first-position gaps of the closed-first group within 4 points; assuming instead that the format makes no difference misses the two fee gaps, net and one-value, by 9 and 14 points. Fitted to the closed-first group, the same model overstates the open-ended fee gaps by 7 and 8 points, as the main model does (Appendix E.4.3).

Seeing the closed-ended question first raises written fee mentions from 17% to 26%. The model captures this through a lower writing threshold for fees after the closed-ended question. For sustainability, written mentions are slightly lower after the closed-ended question (34% instead of 35%). The model cannot produce such a decline, because it

[16] These respondents make up 19% of sustainability writers and 14% of fee writers in the closed-first group, and 8% and 9% in the open-first group. Their gaps exceed those of writers who also select the topic (for sustainability in the closed-first group, 31% against 13%). Excluding them from both the estimation and the raw comparison would remove more respondents from the closed-first group than from the open-first group (6.5% against 2.7% of respondents for sustainability), which would unbalance the randomized groups.

assumes that seeing the list can only make writing easier or leave it unchanged. The estimate therefore sets the sustainability threshold after the list equal to the one without it ($\eta_S = 1$); all fitted reporting shares remain within one percentage point of the observed ones (Appendix Table B.7).

C Descriptive decomposition of the value–action gap

Each gap evaluates choice when the target topic is reported and its competitor is not. For fees, the gap is high-fee choice, y_i; for sustainability, it is conventional-fund choice, $1-y_i$. Table C.1 reports both the raw mean within this reporting group and the fitted gap from Section 5's additive OLS specification. The two calculations use the same reporting definitions, but need not give the same result.

I retain $T_i^{m,\omega,j}$ from Section 5. Here $y_i = 1$ denotes sustainable high-fee fund choice, $m_S, m_F \in \{O, C\}$ identify the sustainability and fee reporting formats, and $\omega \in \{OC, CO\}$ identifies question order. Open reports are mentions; closed reports are selections.

Raw gaps and the across-order comparison. The raw comparison uses first-position open reports at the start and first-position closed reports at the end. It follows rows $1 \to 3 \to 7 \to 8$ in Table C.1. Both topics use the same sequence: change the fee-report format, then question order, then the sustainability-report format.

Let $G_{j,r}^{\text{raw}}$ denote the raw gap for topic j in row r of Table C.1. The fee group $\mathcal{I}_{F,r}$ contains respondents in that row's question-order arm who report fees and not sustainability ($T_i^{m_F,\omega,F} = 1$, $T_i^{m_S,\omega,S} = 0$). The sustainability group $\mathcal{I}_{S,r}$ reverses these settings. With $n_{j,r}$ denoting the corresponding group count shown in brackets in the table, the raw gaps are simply group averages:

$$G_{F,r}^{\text{raw}} = \frac{1}{n_{F,r}} \sum_{i \in \mathcal{I}_{F,r}} y_i, \qquad G_{S,r}^{\text{raw}} = \frac{1}{n_{S,r}} \sum_{i \in \mathcal{I}_{S,r}} (1 - y_i).$$

Multiplying these averages by 100 gives the percentages in the raw columns.

OLS-fitted gaps. For each format combination and question order, I estimate

$$\begin{aligned} y_i = \alpha^{m_S,m_F,\omega} &+ \beta_S^{m_S,m_F,\omega} T_i^{m_S,\omega,S} \\ &+ \beta_F^{m_S,m_F,\omega} T_i^{m_F,\omega,F} + \varepsilon_i^{m_S,m_F,\omega}. \end{aligned}$$

Setting the target indicator to one and the competing indicator to zero gives

$$\begin{aligned} G_F^{m_S,m_F,\omega} &= \widehat{\alpha}^{m_S,m_F,\omega} + \widehat{\beta}_F^{m_S,m_F,\omega}, \\ G_S^{m_S,m_F,\omega} &= 1 - \widehat{\alpha}^{m_S,m_F,\omega} - \widehat{\beta}_S^{m_S,m_F,\omega}. \end{aligned}$$

The format indices always mean *sustainability, then fees.* When $m_S = m_F = m$, this is equation (1) and $G_j^{m,m,\omega} = G_j^{m,\omega}$. Each fit uses everyone in its question-order arm, with no additional controls or choice-card fixed effects.

Following one of the two paths of eq. (11), I change the target topic's format first and the competing topic's format second, holding question order fixed. The intermediate fit uses $(m_S, m_F) = (O, C)$ for fees and (C, O) for sustainability. The examples below show how the three fitted gaps are calculated for the open-first arm ($\omega = OC$).

Initial and final coefficients come from Table 4, columns (1) and (4); intermediate coefficients come from the additional mixed-format fits, which are not displayed in that table. Each line uses its own regression. The superscripts on G identify that fit. The printed coefficients are rounded and therefore need not reproduce every decimal of the fitted gaps. The decomposition examples in the subsections below use rounded gaps; the tabled estimates and intervals use unrounded values.

Example: fees in the open-first arm. The fee gap sets the fee indicator to one and the sustainability indicator to zero. It therefore adds the intercept and the fee coefficient:

$$G_F^{O,O,OC} \simeq \underbrace{0.594}_{\widehat{\alpha}} + \underbrace{0.257}_{\widehat{\beta}_S} \times 0 + \underbrace{(-0.177)}_{\widehat{\beta}_F} \times 1 = 0.417,$$

$$G_F^{O,C,OC} \simeq \underbrace{0.629}_{\widehat{\alpha}} + \underbrace{(-0.125)}_{\widehat{\beta}_F} = 0.503,$$

$$G_F^{C,C,OC} \simeq \underbrace{0.565}_{\widehat{\alpha}} + \underbrace{(-0.134)}_{\widehat{\beta}_F} = 0.431.$$

Example: sustainability in the open-first arm. The sustainability gap sets the sustainability indicator to one and the fee indicator to zero. The intercept plus the sustainability coefficient predicts sustainable-fund choice; subtracting this prediction from one gives the gap:

$$G_S^{O,O,OC} \simeq 1 - \left[\underbrace{0.594}_{\widehat{\alpha}} + \underbrace{0.257}_{\widehat{\beta}_S} \times 1 + \underbrace{(-0.177)}_{\widehat{\beta}_F} \times 0 \right] = 0.149,$$

$$G_S^{C,O,OC} \simeq 1 - \left[\underbrace{0.527}_{\widehat{\alpha}} + \underbrace{0.307}_{\widehat{\beta}_S} \right] = 0.166,$$

$$G_S^{C,C,OC} \simeq 1 - \left[\underbrace{0.565}_{\widehat{\alpha}} + \underbrace{0.316}_{\widehat{\beta}_S} \right] = 0.119.$$

Table C.1: Raw and OLS gaps under the same reporting definitions

	Reporting format		Fee gap (%)		Sustainability gap (%)	
Row	Sustainability	Fees	Raw [n]	OLS (SE)	Raw [n]	OLS (SE)
Open first ($\omega = OC$, $N = 1{,}565$)						
1	Open	Open	39.7 [194]	41.7 (3.1)	15.7 [490]	14.9 (1.6)
2	Closed	Open	32.0 [147]	34.9 (3.1)	17.2 [693]	16.6 (1.4)
3	Open	Closed	49.3 [556]	50.3 (1.9)	12.8 [274]	10.6 (1.9)
4	Closed	Closed	41.1 [409]	43.1 (2.1)	14.1 [377]	11.9 (1.7)
Closed first ($\omega = CO$, $N = 1{,}494$)						
5	Open	Open	50.0 [266]	51.1 (2.6)	14.5 [385]	13.8 (1.7)
6	Closed	Open	45.3 [234]	47.5 (2.6)	14.1 [462]	13.0 (1.6)
7	Open	Closed	54.7 [536]	54.9 (1.9)	11.3 [238]	10.9 (1.9)
8	Closed	Closed	50.5 [463]	50.8 (2.0)	9.9 [274]	9.4 (1.8)

Note: This table shows observed group means and OLS-fitted gaps under the same reporting conditions. Brackets give raw group counts n; each regression uses all N respondents in its arm. Parentheses give HC1 standard errors in percentage points. Calculations use unrounded estimates. From row 1 to row 8, the OLS changes have standard errors of 3.66 points for fees ($p = 0.012$) and 2.40 for sustainability ($p = 0.020$), and the raw changes 4.2 and 2.4 points (two-sample binomial; $p = 0.010$ and 0.016). In the OLS fee change, neither the intercept part (3.4 points, $p = 0.19$) nor the coefficient part (5.7 points, $p = 0.15$) is significant at the 10% level on its own, but their sum is, because both point in the same direction and their estimates are negatively correlated.

C.1 Raw comparison across question-order arms

For either topic, the total change from row 1 to row 8 of Table C.1 is the sum of three successive differences:

$$G^{\text{raw}}_{j,8} - G^{\text{raw}}_{j,1} = \underbrace{\left(G^{\text{raw}}_{j,3} - G^{\text{raw}}_{j,1}\right)}_{\text{fee-report format}} + \underbrace{\left(G^{\text{raw}}_{j,7} - G^{\text{raw}}_{j,3}\right)}_{\text{question-order difference}} + \underbrace{\left(G^{\text{raw}}_{j,8} - G^{\text{raw}}_{j,7}\right)}_{\text{sustainability-report format}}. \tag{19}$$

The intermediate gaps in rows 3 and 7 cancel, leaving the final gap minus the initial gap. The observed percentages illustrate this sequence:

$$\text{Fees:} \quad 39.7\% \xrightarrow{\text{fee report}} 49.3\% \xrightarrow{\text{question order}} 54.7\% \xrightarrow{\text{sust. report}} 50.5\%,$$

$$\text{Sustainability:} \quad 15.7\% \xrightarrow{\text{fee report}} 12.8\% \xrightarrow{\text{question order}} 11.3\% \xrightarrow{\text{sust. report}} 9.9\%.$$

The intermediate groups use open sustainability reports and closed fee reports in both arms. For fees, the target definition changes first; for sustainability, the competing definition changes first. Table C.2, Panel A, reports the successive changes and their uncertainty. The displayed levels and components are rounded separately; the calculation

uses unrounded values. Changing the sequence can change the individual components, but not the overall difference between the endpoints.

These comparisons change reporting groups, not an individual's choice. Even the middle step can reflect changes in group composition: seeing the closed-ended question may affect who later mentions a topic. It is therefore an order *difference*, not an identified correction. Paired answers illustrate the distinction between nonmention and nonselection: among open-first sustainability writers who do not mention fees, only 54.7% also leave fees unselected. This describes agreement between reports, not the absence of a latent fee concern.

Relation to the structural comparison. Appendix B.8 compares these raw changes with the structural model. For the one-value gaps in Table 3, the differences between the raw and model changes, 6.0 and -1.8 points, have 95% intervals of $[-5.4, 11.4]$ and $[-5.9, 2.3]$, computed over the same 2,000 bootstrap resamples as the structural estimates. When the raw data count respondents who write about a topic without selecting it as closed-ended reporters, as the model does, the one-value changes are $+14.4$ and $+1.1$ points, and the differences, 5.9 and 0.2 points, have 95% intervals of $[-5.3, 11.2]$ and $[-3.9, 4.3]$.

At the open endpoint, the model predicts a fee gap of 41.8% among fee writers who do not write about sustainability and 52.4% among all fee writers. The observed gaps are 39.7% and 48.1%, below these values but within the sampling error of the two groups ($n = 194$ and $n = 258$).

C.2 Mixed-format OLS within question order

Using the fitted gaps calculated above, I now separate each within-order change into the target-format change and the competing-format change.

Fees in the open-first arm. Using the fitted gaps to two decimal places, rows $1 \to 3 \to 4$ of Table C.1 give

$$\underbrace{43.07 - 41.69}_{\substack{\text{total}\\ +1.38}} = \underbrace{50.34 - 41.69}_{\substack{\text{fee format}\\ +8.65}} + \underbrace{43.07 - 50.34}_{\substack{\text{sustainability format}\\ -7.27}} . \tag{20a}$$

The fee-format increase is largely offset by changing the competing sustainability format. All differences are in percentage points.

Sustainability in the open-first arm. Using the rounded fitted gaps, rows $1 \to 2 \to 4$ give, in percentage points,

$$\underbrace{11.90 - 14.92}_{\substack{\text{total} \\ \approx -3.0}} = \underbrace{16.56 - 14.92}_{\substack{\text{sustainability format} \\ +1.64}} + \underbrace{11.90 - 16.56}_{\substack{\text{fee format} \\ -4.66}}. \tag{20b}$$

The competing-fee change more than offsets the target-format increase, leaving the gap about 3.0 percentage points lower. The initial increase is imprecisely estimated ($p = 0.170$). In each example the intermediate gap cancels. Reversing the calculation can change the two components, but not their sum; "target first" is not question order.

Results and interpretation. In the closed-first arm, the corresponding paths are $5 \to 7 \to 8$ for fees and $5 \to 6 \to 8$ for sustainability. Table C.2, Panel B, reports all target-first contrasts. The fee components largely offset in both arms. Sustainability's competing-format change is negative in both arms, while its target-format change is small and imprecise—positive in the open-first arm and negative in the closed-first arm. Different signs or significance levels do not by themselves establish a difference between arms.

For open-first sustainability, the same sequence gives raw gaps of $15.7\% \to 17.2\% \to 14.1\%$ and fitted gaps of $14.9\% \to 16.6\% \to 11.9\%$. OLS imposes additivity across reporting cells; adding the interaction between the two report indicators reproduces the raw means. Both calculations use the same data.

Within the open-first partition into unengaged, recognition-only and recalled respondents (the engagement levels U, K and R), which excludes write-only respondents on either topic, a saturated engagement specification does not reject additivity (HC1 Wald $\chi^2(4) = 4.07$, $p = 0.40$). Table C.3 reports the nine raw and fitted cells. The largest discrepancy occurs when both values are recalled, so the test supports the OLS as a compact summary rather than an exact description of every cell.

Table C.2: Changes in value–action gaps and their uncertainty

A. Raw: across arms, rows $1 \to 3 \to 7 \to 8$

Target topic	Fee-report change	Question-order difference	Sustainability-report change	Total change
Fees	+9.6	+5.4	−4.1	+10.8
	(3.0)	(3.0)	(1.3)	(4.3)
Sustainability	−2.9	−1.4	−1.5	−5.9
	(1.6)	(2.8)	(1.8)	(2.5)

B. OLS: target first, within each question-order arm

Question order	Target topic	Target format changed first	Competing format changed second	Total change
Open first	Fees	+8.7	−7.3	+1.4
		(2.5)	(1.0)	(2.7)
Open first	Sustainability	+1.6	−4.7	−3.0
		(1.2)	(1.1)	(1.6)
Closed first	Fees	+3.8	−4.0	−0.2
		(1.9)	(0.9)	(2.1)
Closed first	Sustainability	−0.9	−3.6	−4.5
		(1.4)	(1.2)	(1.8)

Note: This table shows percentage-point changes with standard errors in parentheses. Panel A uses 2,000 bootstrap resamples within question-order arms. Panel B uses the joint HC1 covariance across the four OLS fits, accounting for their common respondents. The panels describe different comparisons. Calculations use unrounded estimates; tests are exploratory, two-sided, and unadjusted for multiple comparisons.

Scope of the comparisons. Changing a reporting definition re-estimates the intercept and both slopes, even when the other report's definition stays fixed. Reports need not be nested, and within-arm comparisons use second-position answers. Thus neither fixing question order nor using OLS isolates causal effects on choice or reproduces the structural comparison. Appendix D.1 reports the separate completion-sensitivity analysis.

Table C.3: Raw and additive fitted choice across engagement cells

	Sustainability engagement		
Fee engagement	U	K	R
U	42.5 / 44.3 [318]	19.8 / 19.4 [121]	11.4 / 9.2 [245]
K	53.1 / 52.4 [258]	26.7 / 27.5 [120]	16.9 / 17.3 [207]
R	70.5 / 67.4 [132]	43.5 / 42.5 [46]	24.1 / 32.4 [54]

Note: Each cell shows raw / additive fitted low-fee choice (%), with the cell size in brackets, for the open-first $U/K/R$ partition without write-only respondents. The additive model has indicators for K and R for both topics; an HC1 Wald test does not reject adding their four interactions ($\chi^2(4) = 4.07$, $p = 0.40$), a specification check rather than proof of exact additivity.

D Survey, Sample, and Measurement

The survey company invited 22,391 panel members; 6,115 participated, and 3,059 eligible respondents completed the questionnaire. The analysis sample consists of adult Swiss residents with a private pension plan. Table D.1 presents the summary statistics for all respondents for the socioeconomic variables. The mean age of the respondents was 49 years, and approximately 53% had a university degree. The sample consisted of slightly more men than women, with 54.7% men. About half of the sample was married, and 16% were pensioners. The average household size was 2.4 people. The monthly household income was CHF 9,193 and the mean household wealth was CHF 341,250 (approximately USD 314,000 at the survey date). Appendix Tables G.1 and G.2 define all background variables.

Table D.1: Summary statistics

Variable	Sample value
% Female	45.3
Age (years)	49.3
% University Degree	52.9
% Pensioner	16.1
% Married	49.6
Household size (persons)	2.4
Monthly household income (CHF)	9,193
Household wealth (CHF)	341,250
% Chose sustainable fund	66.7

Note: Summary statistics for the analysis sample ($N = 3{,}059$ Swiss retail investors with a private pension plan; data collected October–November 2021). Rows prefixed "%" report the share of respondents in percent; other rows report sample means in the units indicated. Monetary values are in Swiss francs (CHF); at the survey date 1 CHF $\approx$ 0.92 USD.

D.1 Break-off attrition and uncertainty

Immediate first-question break-off is higher in the open-first arm, but the direction reverses after the second question: selection into the final outcome sample is 74.4% in the open-first arm and 70.2% in the closed-first arm. The excess selection rate is $(0.744 - 0.702)/0.744 = 5.66\%$ of open-first completers. This overall rate does not determine how many excess-selected respondents belong to a group defined by post-randomization reports. Applying 5.66% within each endpoint cell therefore requires an additional proportional-allocation assumption. Under that assumption, trimming each cell from either end gives ranges of $[8.5, 14.5]$ points for the fee endpoint change and $[-6.8, -0.8]$ for sustainability. A separate regression check trims 5.66% of open-first

fee endorsers, leaving nonendorsers unchanged, and gives $[6.2, 12.2]$ points. These are proportional-trimming sensitivity ranges, not Lee bounds (Lee 2009) for the conditional gaps under monotonic selection alone. They do not establish robustness to arbitrary allocations of the missing respondents.

At the least favorable ends, the fee change is 8.5 points (SE 4.3, $p = 0.05$) and the sustainability change -0.8 points (SE 2.3, $p = 0.73$); Imbens–Manski 95% intervals that cover the whole range (Imbens and Manski 2004) are $[1.3, 21.5]$ for fees and $[-10.9, 3.0]$ for sustainability. The sustainability decrease stays significant at the 5% level unless at least 34% of the excess completers in the open-first sustainability cell chose the conventional fund, 2.2 times the cell's rate of 15.7%. If the excess completers may sit in any cell, the fee increase disappears only if 42 of the about 88 were fee writers who chose the low-fee fund, a group that makes up 7.5% of the open-first sample; erasing both changes requires 74 of the 88, all from two groups that together make up 12% of it. A pre-specified test checks the scenario that would break this result: excess completers who engage little and choose closer to random. Writers with answers shorter than the median choose the conventional fund at 15.8% (shortest quartile: 13.6%), and writers in the less engaged half by the length of their later open answers at 21.2% (95% upper bound 27.0%); the last measure uses answers written after the fund choice. Gender, education, and age cannot identify the excess completers: their implied shares have standard errors of about 0.3.

Table D.2: Survey flow and break-off by randomized question order

Stage	Open first	Closed-ended first
Arm entrants	2,104	2,129
Break-off at first question	219 (10.4%)	84 (3.9%)
Break-off at second question	35 (1.9%)	283 (13.8%)
Completed both elicitations	1,850	1,762
Final analysis sample	1,565	1,494
Post-elicitation attrition	15.4%	15.2%
Final completion rate	74.4%	70.2%

Note: Percentages for second-question break-off use the number reaching that question. Both arms encounter both formats. Immediate first-question break-off is higher under open-first assignment, but total selection into the final outcome sample is higher in that arm because second-question break-off is much larger under closed-first assignment. The proportional-trimming sensitivity in Appendix D.1 uses the final completion rates and trims selected groups in the open-first arm.

Table D.3 relates break-off at the open-ended question to respondent characteristics. Women, respondents without a university degree, and respondents with lower wealth were more likely to drop out.[17] Similarly, Hadler (2025) observed that women and respondents with lower education were more likely to break off a survey in response to an open-ended question. The estimated effects here are small, suggesting little impact on external validity.

[17]For total dropout after the first open-ended question (the table's second column), the signs are similar and the magnitudes larger.

Table D.3: OLS regression on survey attrition

	Attrition Open	Attrition Total
Age	0.001	0.003***
	(0.001)	(0.001)
Female	0.028**	0.062***
	(0.012)	(0.016)
University educ	−0.020*	−0.065***
	(0.012)	(0.017)
Pensioner	0.008	0.048*
	(0.019)	(0.027)
Married	−0.007	−0.017
	(0.014)	(0.020)
HH-size	−0.003	−0.007
	(0.006)	(0.008)
Log income	0.005	0.001
	(0.016)	(0.023)
Log wealth	−0.013**	−0.034***
	(0.006)	(0.008)
Num.Obs.	2862	2862
R2	0.009	0.040

Note: This table shows linear-probability regressions of break-off at the first open-ended question (column 1) and at any point after it (column 2; the survey has three further open-ended questions not analyzed here). Covariates come from a 2024 panel refresh rather than the 2021 wave; respondents who did not report income or wealth are excluded. Classical standard errors in parentheses; no fixed effects. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

The fieldwork tracker, which Table D.2 follows, attributes first-position break-offs to 219 of 2,104 open-first entrants (10.4%) and 84 of 2,129 closed-first entrants (3.9%). The raw response export, which records a break-off only when no first-question answer was saved, instead implies first-position break-offs of 209 of 2,104 and 29 of 2,129, a differential of 8.6 percentage points. The export reconciles exactly with the entrant counts. The relevant selection rates for the choice outcome are final completion rates, which do not depend on this convention and whose ordering reverses.

D.2 Topic allocation and answer quality

To assign topics to the open-ended text responses, I followed the semi-manual approach developed by Wekhof and Houde (2023) and later used by Filippini et al. (2024). This method involves creating a comprehensive dictionary for each topic. The researcher specifies an initial set of topics ex ante (informed by the closed-ended question categories) and, after the clustering step described below, manually selects which clustered words define each topic. To identify the relevant words, Wekhof and Houde (2023) proposed a method to cluster keywords into preliminary groups. Following this approach, I tagged all words in the answers as nouns, verbs, adjectives, or adverbs (i.e., "part of speech

tagging") using the "spacy" library for Python (Honnibal et al. 2020). Then, I created a list of unique words and only selected words with a length of at least four characters.

In the third step of the process, I matched words to a pre-trained word-embedding matrix.[18] This matrix assigns a vector to each word that measures its semantic distance to all other words in the corpus. For instance, the distance between "return" and "investment" is smaller than the distance between "return" and "elephant." In the fourth step, I clustered the words into groups using k-means clustering based on the cosine similarity from the embedding matrix, separately for each part-of-speech category. Lastly, I manually reviewed each cluster of words and selected words for the different topics.

Table D.4 shows the summary statistics for the open-ended responses. The mean number of words was 13.3, with a standard deviation of 11. The text corpus with all answers consisted of 4,746 unique words, out of which 1,238 words define the 18 categories of the final dictionary.[19] The 18 categories are the ten topics of the closed-ended question, four topics that appear only in written answers (ease of use, past performance [`development`], fund holdings [`content_fund`], and greenwashing), a "don't know" category, and three markers that flag the literal words "Rendite," "Risiko," and "Nachhaltigkeit" [`wording_sust`].

In addition, 502 words could not be linked to the word embedding, either because they were uncommon or because of spelling mistakes. These words usually occurred only once or twice in the corpus, and where possible I classified them manually.

Following the topic allocation, 198 respondents could not be allocated to any topic. I manually classified these responses. Many of those answers consisted of either "no interest in investing" or "no answer," such as a dash or dots. This classification adds the categories "no interest" and "no answer," so the data contain the 20 categories that Appendix D.3 codes.

Table D.5 shows the ten most important words for the topics "sustainability" and "fees."

Poor-quality responses (often called "non-response") typically include answers such as "do not know," "no interest," or random characters (Geer 1988, Hadler 2025).[20] In this study, the non-response rate for the sub-sample answering the open-ended question first was 6.7% (components rounded): 2.9% wrote that they were not interested in the topic, 2.4% wrote a form of "do not know," and 1.3% gave no answer or random characters.

[18] I used the fastText word embeddings of Grave et al. (2018) for German.

[19] All words had at least four characters, except for the following abbreviations that were added manually "ESG," "ETF," and "TER" (for total expense ratio).

[20] Blank answers were technically not possible in this survey.

Table D.4: Summary statistics for open-ended responses

Metric	Value
# answers	3,059
mean # words	13.3
sd # words	11.0
mean # sentences	1.6
sd # sentences	0.9
total # unique words	4,746
total # words used for topics	1,238

Note: This table shows statistics of the open-ended answers on the most important factors in choosing a fund (N = 3,059). Means and standard deviations are across respondents; the unique-word rows are corpus totals.

Table D.5: Most frequent words for the topics sustainability and fees

sustainability		*fees*	
word	freq.	word	freq.
sustainability	423	fees	246
sustainable	229	costs	171
sustainable (plural)	134	administrative costs	54
ecological	35	administration fees	51
ethical	31	expenses	25
social	30	deep	23
sustainable (acc.)	29	low	9
ecological	20	favorable	7
environmentally friendly	12	deposit fees	6
social	11	accrued	6

Note: The most frequent words for the topics sustainability and fees. The words were originally in German and translated for this table. The word "sustainable" appears three times because German distinguishes its singular, plural, and accusative forms. "Deep" is the literal translation of "tief," which in Swiss usage means *low* (as in "tiefe Gebühren," low fees).

Non-responses in this study's results also align with the lower bound from the literature: Hadler (2025) reported non-response rates between 12.2% and 21.6%. Behr et al. (2014) found non-response rates for open-ended probes to be between 10.1% and 12.6%, while Meitinger and Behr (2016) found rates between 10.5% and 31.4%. Geer (1988) observed non-response rates of about 5% in multiple surveys and suggested that lack of interest, rather than respondents' rhetorical abilities, was the main reason for non-response. Geer (1988, 1991) also noted that whenever respondents provided open-ended answers, they were generally well-written, reflective, and coherent, which aligns with my observations on this corpus.

D.3 Application-specific validation of the keyword dictionary

The dictionary method used in this paper was originally developed by Wekhof and Houde (2023), who validated it on a different corpus (German respondents discussing energy-efficiency retrofits). In the validation of Wekhof and Houde (2023), the method's agreement with two independent human coders approximated the inter-coder agreement. The keywords chosen for the present application (Swiss retail investors describing fund-selection criteria) have not previously been independently validated. This subsection provides application-specific validation against two reference sources: (i) hand-codes I produced on a random subsample, and (ii) a large-language-model coder applied to the full corpus.

Hand-coding validation. I drew a simple-random sample of 500 open-ended responses from the 3,059 non-empty responses and hand-coded each for the two load-bearing topics: fees and sustainability. Hand-coding followed a written codebook that documents the inclusion definition, illustrative German seed keywords, and edge cases for each topic. The single-coder hand-codes serve as the author-coded reference labels for these two topics; the remaining 18 topics were not hand-coded because the paper's claims do not rest on them.

LLM cross-check. All 3,059 responses were independently coded for all 20 topics by OpenAI's GPT-5.4 (model identifier `gpt-5.4`; coding run of May 24, 2026) using the same codebook supplied as a system prompt with structured-JSON output (temperature zero; one API call per response so each is coded in isolation). The LLM serves two roles. First, on the 500-response calibration sample it is a second rater on fees and sustainability, benchmarked against the hand-codes alongside the keyword dictionary. Second, on the full 3,059-response corpus it serves as a robustness check on every topic, including those for which hand-codes are unavailable. Two caveats about the design of this validation: (i) the hand-codes used as the gold-standard reference were produced by the author, who also constructed the keyword dictionary, so this coder–instrument dependence would be resolved by an independent second hand-coder, but for the two extremely concrete

topics here (any mention of "sustainability" or "fees"; the topic boundaries are objective) the bias from this dependence is likely small; (ii) because the LLM was prompted with the same codebook the keyword dictionary embodies, the LLM-vs-keyword agreement primarily measures consistency in following the codebook rather than fully independent validation. The author-coded labels provide a direct reference for the two main topics, but they are not independent of dictionary construction, and the LLM check is not fully independent because it applies the same codebook. Together, the two checks establish that the reported results are robust to two implementations of the coding rule; they do not substitute for an independent human validation. The LLM additionally extends the codebook-following check to the full sample.

Calibration results. Table D.6, Panel A reports per-topic precision, recall, F_1, and Cohen's κ against the hand-coded reference, with 95% bias-corrected accelerated bootstrap confidence intervals ($B = 1{,}000$). For both main topics, the keyword dictionary achieves $\kappa \geq 0.95$ with precision and recall above 95%. The LLM achieves comparable agreement ($\kappa \geq 0.97$). Bootstrap intervals heavily overlap, so the two classifiers are statistically indistinguishable on the topics that carry the paper's claims.

Full-sample concordance. On the full 3,059-response corpus across all 20 topics, the keyword dictionary and the LLM agree on 97% of classifications at the median topic; the two main topics agree on 98.3% of responses ($\kappa = 0.95$ for fees, 0.96 for sustainability). Three topics show lower concordance: `wording_sust` (86% agreement, $\kappa = 0.60$), `content_fund` (85%, $\kappa = 0.59$), and `development` (96%, $\kappa = 0.60$). The first is a literal-word marker (any mention of "Nachhaltigkeit") not used in the main analysis; the second and third are broad, vaguely-defined categories where the keyword dictionary applies a narrower inclusion rule than the LLM. None of these topics enter the structural estimation or the main regressions.

Robustness of the gap estimates. Table D.6, Panel B substitutes the LLM open-ended labels for the keyword labels in column (1) of Table 4 and re-predicts the value–action gap at the same covariate profile used in the main text. For fees, the predicted gap moves from 41.7% (keyword) to 40.8% (LLM), a difference of -0.9 percentage points within bootstrap-CI overlap. For sustainability, the predicted gap moves from 14.9% to 14.2% (-0.7 pp, also within overlap). The paper's value–action-gap numbers are robust to the choice of open-ended coder. This substitution is a reduced-form coding check: it recomputes the OLS-based gaps but does not re-estimate the structural reporting thresholds, writer groups, or choice model under the alternative labels.

Table D.6: Validation of the keyword dictionary for the fees and sustainability topics

	n_{pos}	Precision	Recall	F_1	Cohen's κ
Panel A. Calibration validation against hand-codes (n=500)					
Keyword dictionary					
Fees	85	95.3 [87, 99]	96.5 [91, 99]	95.9 [92, 98]	0.95 [.90, .98]
Sustainability	178	96.7 [93, 99]	98.9 [97, 100]	97.8 [96, 99]	0.97 [.93, .98]
LLM (GPT-5.4)					
Fees	85	97.6 [93, 100]	97.6 [93, 100]	97.6 [94, 99]	0.97 [.93, .99]
Sustainability	178	98.9 [96, 100]	98.3 [96, 100]	98.6 [97, 100]	0.98 [.95, .99]

Panel B. Robustness of the value–action gap to coder choice			
	Predicted gap [95% CI]		Diff
Specification	Keyword	LLM	(pp)
Fees, open-ended (col 1)	41.7 [35.6, 47.4]	40.8 [35.0, 47.4]	−0.9
Sustainability, open-ended (col 1)	14.9 [11.9, 18.2]	14.2 [11.2, 17.3]	−0.7

Note: Panel A shows precision, recall, and F_1 (in percent) and Cohen's κ against single-coder hand codes of 500 randomly drawn responses, with 95% bias-corrected accelerated bootstrap intervals in brackets ($B = 1{,}000$). The LLM (GPT-5.4, temperature zero) coded each response separately with the same codebook. Panel B shows the value–action gap (%) from Table 4, column (1), with keyword or LLM labels for the open-ended answers; closed-ended answers are unchanged.

D.4 Question wording

In both question orders, the closed-ended options were displayed in randomized order. Respondents who started with the open-ended question received the following text:

> *Financial decisions are often complex and vary from person to person. We would like to ask you to answer some questions in writing. Your individual text response will help us better understand how you come to a decision.*
>
> *Please imagine the following situation: You received CHF 10,000 and would like to invest this sum over a period of 10 years. Your bank offers you a broad selection of mutual funds.*
>
> *What criteria would be important to you when choosing a fund? Please write a short text with about three sentences.*

I presented respondents who first answered the open-ended question with the following closed-ended question:

> *Please imagine the following situation again: You received CHF 10,000 and would like to invest this sum over a period of 10 years. Your bank offers you a broad selection of mutual funds.*
> *What criteria would be important to you when choosing a fund? (multiple options possible).*
>
> ☐ *Expected return*
>
> ☐ *Expected Risk*
>
> ☐ *Management fees*

☐ *Transparency of product information*

☐ *Sustainability of the included Companies*

☐ *Control over the Investment*

☐ *Recommendation of the bank advisor*

☐ *Trust in the bank*

☐ *Advice from third parties (friends, family, etc.)*

☐ *Actively or passively managed fund (e.g., ETF)*

☐ *Other (please specify)* [text field][21]

[21] Out of 3,059 respondents, 112 used the "other" option. Twelve respondents mentioned sustainability-related topics, and three mentioned fees. I did not use the "other" responses for the analysis.

Respondents who started with the closed-ended question were presented with the following text:

> *Please imagine the following situation: You received CHF 10,000 and would like to invest this sum over a period of 10 years. Your bank offers you a broad selection of mutual funds.*
> *What criteria would be important to you when choosing a fund? (multiple options possible).*

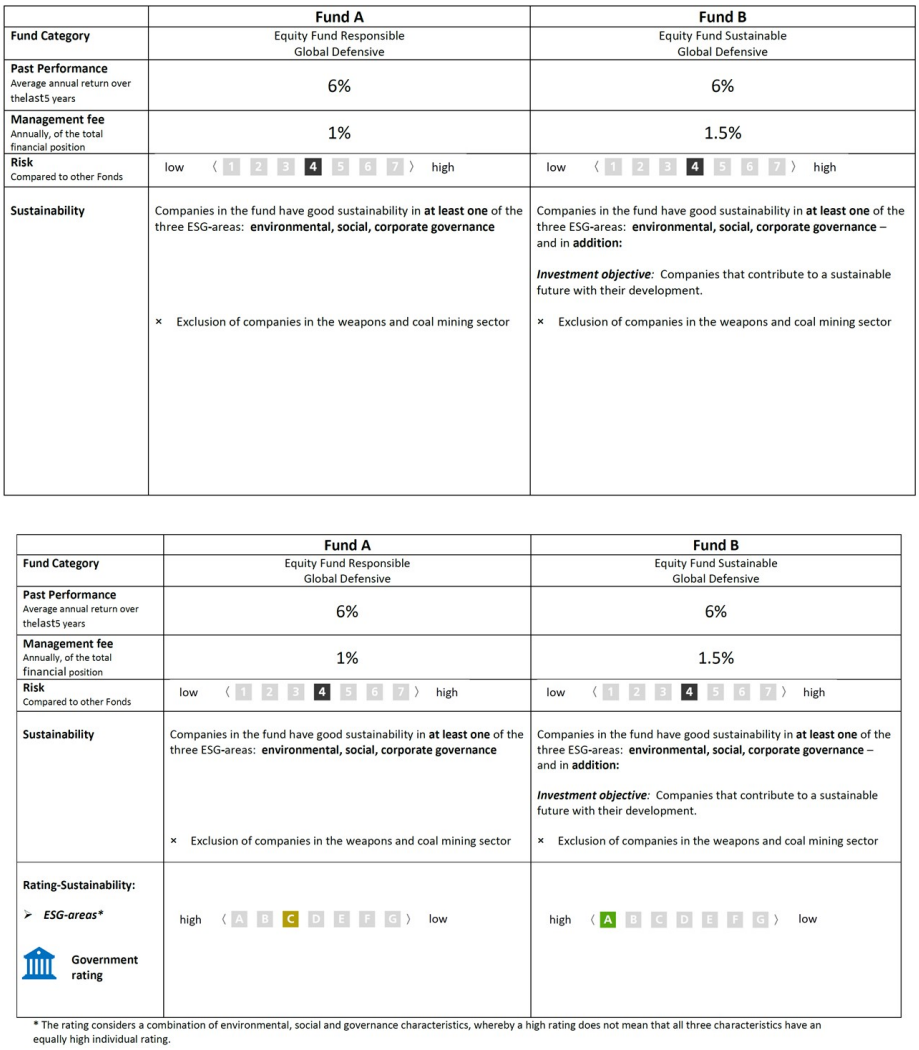
- ☐ *Expected return*
- ☐ *Expected Risk*
- ☐ *Management fees*
- ☐ *Transparency of product information*
- ☐ *Sustainability of the included Companies*
- ☐ *Control over the Investment*
- ☐ *Recommendation of the bank advisor*
- ☐ *Trust in the bank*
- ☐ *Advice from third parties (friends, family, etc.)*
- ☐ *Actively or passively managed fund (e.g., ETF)*
- ☐ *Other (please specify)* [this option included a free text field]

After answering the closed-ended question, these respondents then were presented with the open-ended question:

> *Financial decisions are often complex and vary from person to person. We would like to ask you to answer some questions in writing. Your individual text response will help us better understand how you come to a decision.*
>
> *Please imagine again the situation of the previous question: You received CHF 10,000 and would like to invest this sum over a period of 10 years. Your bank offers you a broad selection of mutual funds.*
>
> *What criteria would be important to you when choosing a fund? Please write a short text with about 3 sentences.*

D.5 Choice cards

Figure D.1: Randomized fund-choice cards: control and public rating



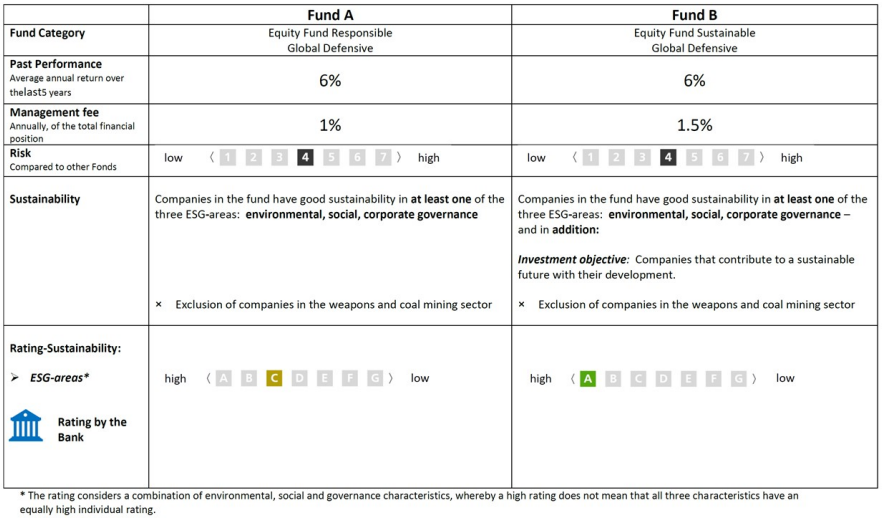
	Fund A	Fund B
Fund Category	Equity Fund Responsible Global Defensive	Equity Fund Sustainable Global Defensive
Past Performance Average annual return over thelast5 years	6%	6%
Management fee Annually, of the total financial position	1%	1.5%
Risk Compared to other Fonds	low 〈 1 2 3 **4** 5 6 7 〉 high	low 〈 1 2 3 **4** 5 6 7 〉 high
Sustainability	Companies in the fund have good sustainability in **at least one** of the three ESG-areas: **environmental, social, corporate governance** × Exclusion of companies in the weapons and coal mining sector	Companies in the fund have good sustainability in **at least one** of the three ESG-areas: **environmental, social, corporate governance** – and in **addition:** ***Investment objective***: Companies that contribute to a sustainable future with their development. × Exclusion of companies in the weapons and coal mining sector

	Fund A	Fund B
Fund Category	Equity Fund Responsible Global Defensive	Equity Fund Sustainable Global Defensive
Past Performance Average annual return over thelast5 years	6%	6%
Management fee Annually, of the total financial position	1%	1.5%
Risk Compared to other Fonds	low 〈 1 2 3 **4** 5 6 7 〉 high	low 〈 1 2 3 **4** 5 6 7 〉 high
Sustainability	Companies in the fund have good sustainability in **at least one** of the three ESG-areas: **environmental, social, corporate governance** × Exclusion of companies in the weapons and coal mining sector	Companies in the fund have good sustainability in **at least one** of the three ESG-areas: **environmental, social, corporate governance** – and in **addition:** ***Investment objective***: Companies that contribute to a sustainable future with their development. × Exclusion of companies in the weapons and coal mining sector
Rating-Sustainability: ➢ ***ESG-areas**** **Government rating**	high 〈 A B **C** D E F G 〉 low	high 〈 **A** B C D E F G 〉 low

* The rating considers a combination of environmental, social and governance characteristics, whereby a high rating does not mean that all three characteristics have an equally high individual rating.

Note: Control card (top) and government-issued ESG rating (bottom). The underlying fund attributes are constant.

Figure D.2: Randomized fund-choice cards: bank and climate ratings



	Fund A	Fund B
Fund Category	Equity Fund Responsible Global Defensive	Equity Fund Sustainable Global Defensive
Past Performance Average annual return over thelast5 years	6%	6%
Management fee Annually, of the total financial position	1%	1.5%
Risk Compared to other Fonds	low 〈 1 2 3 **4** 5 6 7 〉 high	low 〈 1 2 3 **4** 5 6 7 〉 high
Sustainability	Companies in the fund have good sustainability in **at least one** of the three ESG-areas: **environmental, social, corporate governance** × Exclusion of companies in the weapons and coal mining sector	Companies in the fund have good sustainability in **at least one** of the three ESG-areas: **environmental, social, corporate governance** – and in **addition:** ***Investment objective***: Companies that contribute to a sustainable future with their development. × Exclusion of companies in the weapons and coal mining sector
Rating-Sustainability: ➢ ***ESG-areas**** **Rating by the Bank**	high 〈 A B **C** D E F G 〉 low	high 〈 **A** B C D E F G 〉 low

* The rating considers a combination of environmental, social and governance characteristics, whereby a high rating does not mean that all three characteristics have an equally high individual rating.

	Fund A	Fund B
Fund Category	Equity Fund Responsible Global Defensive	Equity Fund Sustainable Global Defensive
Past Performance Average annual return over thelast5 years	6%	6%
Management fee Annually, of the total financial position	1%	1.5%
Risk Compared to other Fonds	low 〈 1 2 3 **4** 5 6 7 〉 high	low 〈 1 2 3 **4** 5 6 7 〉 high
Sustainability	Companies in the fund have good sustainability in **at least one** of the three ESG-areas: **environmental, social, corporate governance** × Exclusion of companies in the weapons and coal mining sector	Companies in the fund have good sustainability in **at least one** of the three ESG-areas: **environmental, social, corporate governance** – and in **addition:** ***Investment objective***: Companies that contribute to a sustainable future with their development. × Exclusion of companies in the weapons and coal mining sector
Rating-Sustainability: ➢ ***Climate mitigation*** **Government rating**	high 〈 A B **C** D E F G 〉 low	high 〈 **A** B C D E F G 〉 low

Note: ESG rating issued by the bank (top) and government rating covering climate mitigation only (bottom). Together with the government-rating card, the treatments vary the presence, issuer, and coverage of the ESG label.

E Broader Descriptive Evidence

This appendix places the two focal topics in the full response distribution and documents the ranking reversal and order effects.

In the closed-ended answers (Table E.1), *expected return*, *risk*, and *trust in the bank* were the most frequently selected options; between 57% and 59% of respondents chose these elements, followed by *fees*, *transparency*, and *sustainability* with frequencies ranging from 41% to 54%. Topics with a lower frequency were *consulting bank advisors*, *having control over the assets*, *choosing ETFs* (Exchange-Traded Funds), and *seeking external advice*. The six most frequent topics had similar prevalence across the sample.

Compared to the closed-ended topics, the open-ended counterparts showed lower frequencies almost everywhere, with the gap widening down the list. Only *risk* appeared with a statistically similar frequency in both question types; *sustainability*'s difference was moderate (5.6 percentage points) and *return*'s substantial (19.1), while the *fees* topic was mentioned three times less often in the open-ended answer (16.5%) and *trust in the bank* nine times less (6.3%). Compared to the closed-ended responses, the open-ended answers contained more diverse topics with smaller frequencies and, importantly, topics not available in the closed-ended responses. The frequency difference was statistically significant for all topics at the 1% level, except for *risk*.

The topic rankings differed between the two question types mainly in the middle of the list, with the extremes more stable, which aligns with previous studies (Reja et al. 2003, Schuman et al. 1986). *Risk* and *return* lead the open-ended ranking and remain in the closed-ended's top group, and *external advice* and *ETFs* stay near the bottom under both formats. The middle re-orders: *sustainability* is ranked higher in the open-ended responses (though with a lower absolute frequency), while the closed-ended question pulls *trust in the bank* and *fees* into a top four whose members are statistically indistinguishable from their neighbors and reverses the *risk–return* pair in the estimates only. Appendix E.1 quantifies these reversals across all topic pairs. While many topics occurred with a similar frequency in the closed-ended responses, the relative ranking in the open-ended responses is more distinct, with a larger distance between the topic frequencies.

Table E.1 reports the full order census. Closed-first assignment raises later open reporting for return by 6.7 points, fees by 9.3, transparency by 12.2, control by 3.1, trust in the bank by 10.6, bank-advisor advice by 9.0, and external advice by 2.1; each is significant at the 1% level. It does not significantly change later open risk, sustainability, or ETF reporting. The rise in average open topics from 2.1 to 2.6 is therefore broad but not uniform across topics.

Table E.1: Topic shares in open- and closed-ended answers, separated by ordering

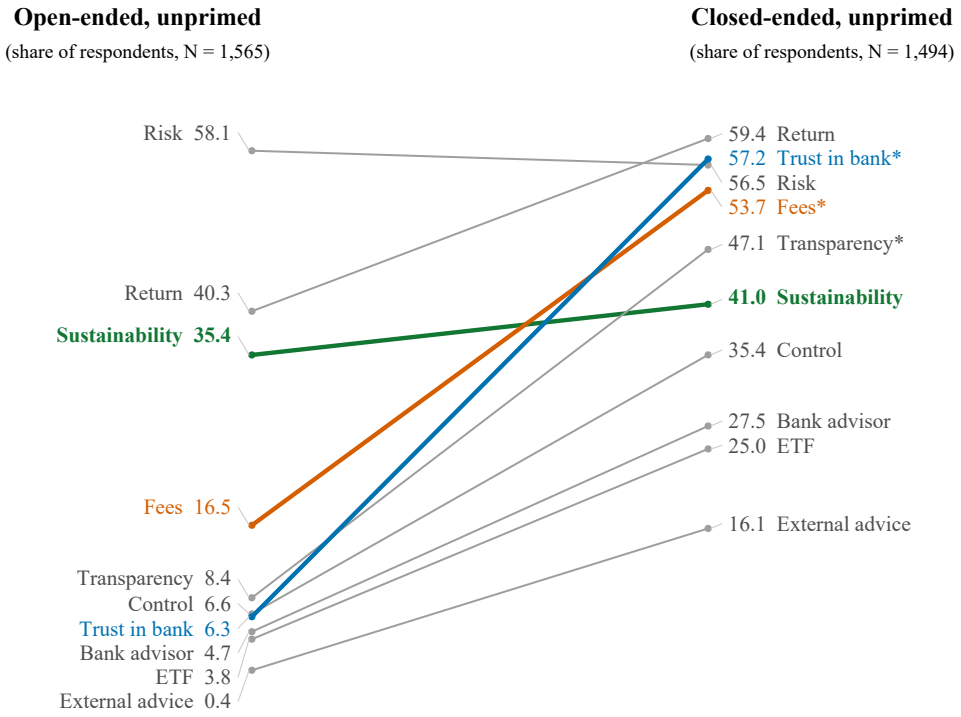
Topic	*Open-ended*		*Closed-ended*		Difference
	First	Second	First	Second	(first)
Risk	58.1	56.4	56.5	62.5	−1.6
Return	40.3	46.9	59.4	62.7	19.1***
Sustainability	35.4	33.7	41.0	51.4	5.6***
Content	20.4	19.5			
Fees	16.5	25.8	53.7	53.4	37.2***
Transparent	8.4	20.6	47.1	55.3	38.7***
Control	6.6	9.7	35.4	33.5	28.8***
Trust Bank	6.3	16.9	57.2	56.8	50.9***
Bank Advisor	4.7	13.7	27.5	27.3	22.8***
Development	4.3	4.1			
ETF	3.8	4.9	25.0	24.7	21.2***
Hassle	2.1	2.7			
Advice External	0.4	2.5	16.1	13.9	15.7***
Greenwashing	0.4	0.2			
Non-Response					
No Interest	2.9	1.5			
Do Not Know	2.4	1.7			
No Answer	1.3	2.4			
# Topics	2.1	2.6	4.2	4.4	2.1***

Note: This table shows the share of respondents (%) who report each topic and the mean number of topics, by format and position: *First* for respondents who answered that format first, *Second* for those who answered it after the other format. Blank cells mark topics not offered on the closed-ended question. *Difference* is closed-ended minus open-ended among first answers, with two-sided two-sample tests: $^{***}p < 0.01$.

Writing first also changes some later closed-ended responses. Risk rises by 6.0 points, sustainability by 10.3, and transparency by 8.1, each significant at the 1% level; return rises by 3.3 points and external advice falls by 2.2, both significant at the 10% level. The remaining closed-ended changes are not significant. These are net order effects: they can combine position, deliberation, consistency, fatigue, and cross-format priming. Standard errors and significance levels come from 2,000 arm-stratified bootstrap draws.

The descriptive results underscore the main format differences. Open answers have lower topic frequencies and a clearer ranking, while the closed-ended compresses the top into a statistically indistinguishable group and reorders the middle (Appendix E.1). Open answers also admit topics absent from the closed-ended question. Order changes several reports, but not by one generic shift shared across topics.

Figure E.1: Topic rankings under open-ended and closed-ended elicitation



Note: Unprimed shares by randomization arm: open-ended answers from the open-first arm ($N = 1{,}565$) and closed-ended answers from the closed-first arm ($N = 1{,}494$). Starred topics overtake sustainability on the closed-ended question; all three reversals survive a Bonferroni correction across the 45 matched topic pairs.

E.1 The ranking flip across all matched topics

This subsection treats the ranking instability systematically. Ten topics were elicited in both formats. Comparing the unprimed measurements, open-ended answers from the open-first arm (N = 1,565) and closed-ended answers from the closed-first arm (N = 1,494), gives 45 topic pairs, and for each pair two within-arm ordering tests: the paired difference of the two topics' mention shares under the open-ended format, and the same difference under the closed-ended format. A pair *reverses order* when the two differences have opposite signs; a reversal is *strict* when both legs are individually significant ($|z|>$ 1.96), and Bonferroni-robust when both clear the 0.05/45 threshold. Standard errors come from a nonparametric bootstrap stratified by arm (B = 2,000), which also delivers the joint covariance used below.

Eight of the 45 pairs reverse order in the estimates, four strictly, and three survive the Bonferroni correction: sustainability against fees ($z = +11.4$ open-ended, -7.1 closed-ended), against transparency ($+18.9$, -3.7), and against trust in the bank ($+20.5$, -9.0). Sustainability falls from third in the open-ended ranking, below only risk and return (weakest pairwise link $z = 2.7$), to sixth on the closed-ended question, below all five topics above it ($z \geq 3.7$) and above control ($z = 3.0$). Trust in the bank moves the other way: mentioned spontaneously by 6.3 percent of respondents, statistically tied with control near the bottom of the open-ended list, it is endorsed by 57.2 percent when offered, joining the closed-ended's top group. The risk and return pair reverses in the estimates only (risk leads by 17.8 points open-ended, $z = +11.4$; return by 2.9 on the closed-ended question, $z = -1.9$). The reversals are not an artifact of comparing different respondents across arms: rerunning the census fully within person gives eight reversals with two strict in the open-first arm (spontaneous mentions against the same respondents' subsequent closed-ended selections) and eight with five strict in the closed-first arm.

The two rankings also differ in how much order they carry. The open-ended ranking is nearly a strict chain: every adjacent pair of shares separates at conventional levels except control and trust in the bank ($z = 0.4$) and bank advisor and ETF ($z = 1.1$). The closed-ended ranking compresses: its top four (return, trust in the bank, risk, fees) sit within 5.7 percentage points of one another with no significant adjacent link, and its ten shares span 16 to 59 percent against the open-ended column's 0.4 to 58. Rank positions inside closed-ended tie groups are therefore not identified; the pairwise reversals above are the statements the data support. The compression is the ordinal face of the lower closed-ended threshold: it pulls weakly differentiated topics into one endorsement band.

The reversal pattern rejects a natural null. If the closed-ended changed every topic's reporting threshold by a common additive shift on a shared logit or probit scale, the format change would be a monotone transformation of topic-level difficulty and every

ranking would be preserved. This additive structure is exactly the model of aggregate brand-awareness scores of Laurent et al. (1995), in which unaided and aided awareness reflect one latent salience read through format-specific difficulty parameters. Fitting the common shift by generalized least squares with the full bootstrap covariance of the ten per-topic shifts gives $Q = 848.1$ on 9 degrees of freedom on the logit scale ($Q = 865.9$ on the probit scale), rejecting the common-shift null at any conventional level. The per-topic shifts on the probit scale are the prompting premia of Table E.2 expressed on the threshold scale of Section 6.

Table E.2: Unprimed topic shares by arm and prompting premia

Topic	Open first r_{jO}	Closed first r_{jC}	Premium w (pp)	Probit shift
Trust in bank	6.3	57.2	50.8 (1.4)	1.71 (0.06)
Transparency	8.4	47.1	38.7 (1.5)	1.30 (0.06)
Fees	16.5	53.7	37.2 (1.6)	1.07 (0.05)
Control	6.6	35.4	28.8 (1.4)	1.13 (0.06)
Bank advisor	4.7	27.5	22.8 (1.3)	1.08 (0.07)
ETF	3.8	25.0	21.1 (1.2)	1.09 (0.07)
Return	40.3	59.4	19.2 (1.8)	0.49 (0.05)
External advice	0.4	16.1	15.7 (1.0)	1.68 (0.17)
Sustainability	35.4	41.0	5.6 (1.8)	0.15 (0.05)
Risk	58.1	56.5	−1.6 (1.8)	−0.04 (0.05)

Note: Unprimed shares by randomization arm (open-first $N = 1{,}565$, closed-first $N = 1{,}494$), in percent. The premium w is the closed-ended share minus the open-ended share; the probit shift is $\Phi^{-1}(r_{jC}) - \Phi^{-1}(r_{jO})$, the same quantity on the threshold scale, with r_{jm} the share of respondents who report topic j in format m (Section 6.3). Bootstrap standard errors in parentheses (stratified by arm, $B = 2{,}000$). w uses unrounded shares, so it can differ in the last digit from the Difference column of Table E.1, which uses the displayed shares.

Table E.2 reports each topic's unprimed shares and its prompting premium w_j, the additional share of the population expressing the topic under the closed-ended question relative to unprimed open-ended elicitation. The premium ranges from 50.8 percentage points for trust in the bank to -1.6 ($z = -0.9$) for risk: the risk topic, not priced on the choice card and already spontaneously accessible to most respondents, pays no premium and anchors the column as a placebo. The premium is not simply inverse salience: external advice has the rarest spontaneous mention yet a premium below return's.

E.2 Cross-format equivalence across topics

Table E.3: Equivalent answering by topic

Topic	Open: equivalent share		Closed-ended: equivalent share	
	First	Second	First	Second
Return	0.83	0.75	0.60	0.53
Risk	0.73	0.65	0.65	0.68
Fees	0.91	0.86	0.41	0.28
Transparent	0.80	0.71	0.31	0.12
Sustainability	0.92	0.81	0.66	0.64
Control	0.50	0.58	0.16	0.10
Bank Advisor	0.60	0.65	0.32	0.10
Trust Bank	0.76	0.78	0.23	0.08
Advice External	0.67	0.78	0.12	0.02
ETF	0.85	0.88	0.17	0.13

Note: Open equivalence is the share of open mentions also selected on the closed-ended question. Closed-ended equivalence is the share of closed-ended selections also mentioned openly. "First" and "Second" refer to the position of the reference format. The arms contain 1,565 open-first and 1,494 closed-first respondents.

Overall, the open-ended responses were more often equivalent than the closed-ended answers for both sequential orderings. This indicates that when respondents wrote about a topic, they also tended to select it among the closed-ended options. The reverse implication was weaker: many closed-ended selections had no matching open mention, especially for low-accessibility topics such as fees.

E.3 Evidence from another survey

Wave 3 of Andre et al. (2022) elicits the same kind of information in a different domain. Each respondent read one of four hypothetical scenarios in which a variable rises: the oil price, the federal funds rate, income taxes, or government spending. After a baseline scenario, respondents predicted the change in inflation on a nine-point scale (grouped here into higher, unchanged, or lower), explained in a text box how they came up with the prediction, and, after the scenario, ticked the thoughts they had in mind on a list of propagation channels. The authors hand-coded the written answers into mechanism groups and grouped the list items in the same way. I use their cleaned data and coding for the 1,607 households in the oil-price, interest-rate, and income-tax scenarios; the government-spending scenario is left out because its cost group combines channels with opposite implications for inflation. In the other three scenarios, the cost group (firms pass on higher costs) implies rising inflation, and the demand group (households and firms spend less) implies falling inflation. The gap is the share of a channel's reporters whose inflation forecast does not move in the channel's direction; net of the other channel, it uses only reporters who do not report the other channel in the same format.

Table E.4 reports the raw comparison. In every scenario, both channels are written by few respondents (6 to 22 percent) and ticked by many (47 to 72 percent). For the cost channel, the gap is higher among tickers than among writers in all three scenarios, both one channel at a time and net of the demand channel. For the demand channel, one channel at a time, the gap falls for the oil price (6.9 points) and changes little in the other two scenarios; net of the cost channel, it falls by 19.4 points for the oil price and by 7.4 points for the interest rate, and it does not change for income taxes. In the interest-rate scenario, the list adds slightly more respondents per writer to the demand channel than to the cost channel (5.2 against 4.1), the reverse of the other two scenarios and of the pooled shares. Net of the other channel, the cost channel's gap still rises there and the demand channel's still falls, so this scenario does not follow the salience pattern of Section 7.3; neither change is significant at the 5% level.

Table E.4: Written and ticked channels and inflation forecasts in Andre et al. (2022)

	Reports (%)		Gap, one channel (%)			Gap, net of the other (%)		
Scenario	Written	Ticked	Written	Ticked	Change	Written	Ticked	Change
Panel A: Cost channel (inflation should rise)								
Oil price	21.9	71.5	4.1	11.8	7.7	4.5	11.7	7.3
					(2.0)			(3.4)
Interest rate	10.6	53.8	16.4	19.6	3.3	11.9	16.2	4.3
					(4.8)			(6.4)
Income tax	6.4	47.2	5.9	18.4	12.5	5.0	15.0	10.0
					(4.5)			(5.4)
Pooled	13.1	57.7	8.7	16.5	7.9	7.1	14.3	7.2
					(2.3)			(3.0)
Panel B: Demand channel (inflation should fall)								
Oil price	20.1	54.8	96.4	89.5	−6.9	94.7	75.4	−19.4
					(2.2)			(5.7)
Interest rate	10.2	62.9	77.4	80.7	3.4	77.5	70.1	−7.4
					(5.5)			(7.1)
Income tax	17.2	51.1	79.1	82.3	3.2	75.3	76.9	1.5
					(3.7)			(4.4)
Pooled	15.9	56.2	84.5	84.3	−0.3	82.8	74.1	−8.6
					(2.3)			(3.4)

Note: This table uses wave 3 of Andre et al. (2022): 1,607 US households, each predicting the direction of inflation in one scenario. Written reports are the authors' hand-coded mechanism groups in the open-text answers; ticked reports are their groups of list items. The gap is the share of a channel's reporters whose forecast does not move in the channel's direction; net of the other channel, it uses reporters who do not report the other channel in the same format. Changes are ticked minus written, for the same respondents. Pooled values are averages over the scenarios weighted by sample size. Standard errors in parentheses are standard deviations over 2,000 bootstrap resamples of respondents within scenarios. Written groups contain 20 to 122 respondents per scenario.

The model of Section 6 carries over with the forecast in place of the fund choice. In each scenario, the cost and demand channels take the roles of the two values; as in

the main estimator, respondents who write about a channel without ticking it count as writers. The forecast is a conditional logit in the two convictions, as the fund choice in Section 6.2: relative to an unchanged forecast, the utility of a rising forecast increases with the index of cost conviction and that of a falling forecast with the index of demand conviction, and the product of the two indices enters both:

$$\Pr(\text{rise}) : \Pr(\text{same}) : \Pr(\text{fall}) = e^{a_R + \kappa_C h_C + \kappa_{CD}^R h_C h_D} : 1 : e^{a_F + \kappa_D h_D + \kappa_{CD}^F h_C h_D},$$

where $h_j = h_j(Z_j)$ is the index of eq. (4), with one skew per channel. The skews and the two product coefficients are shared across the scenarios. There are no choice cards, no earliness information, and no question-order randomization. I estimate the model by maximum likelihood on the 27 combinations of the two reporting levels and the three forecasts in each scenario, jointly for the three scenarios. Because respondents who report one channel also report the other more often than independence predicts, the specification allows a Gaussian copula between the two ranks (equation (16)); the estimated correlations are 0.16 to 0.39, against -0.07 between the two values in this paper's survey (Appendix B.3). The product terms matter here: without them, the fit is clearly worse ($p < 0.001$), and with them, respondents who hold both channels predict a change in either direction less often. The skews are not jointly significant ($p = 0.09$). A multinomial logit with both convictions in both utilities fits about as well ($p = 0.35$) but turns the screen of the cost channel into a small rise, against the measured fall (Appendix E.4.2).

Table E.5 reports the decomposition. For the cost channel, broadening dominates in every scenario (6.0 to 8.3 points), and the stricter screen of the demand channel changes the gap by at most 3.5 points. For the demand channel, the stricter screen of the cost channel lowers the gap by 9.7 to 16.6 points and dominates in every scenario, also in the interest-rate scenario, where the salience order reverses. Pooled, the split is 7.5 and -1.3 points for the cost channel and 3.7 and -12.3 points for the demand channel, with the same pattern as the split for fees (9.4 and -2.6) and sustainability (0.7 and -2.4) in Section 7.2. The asymmetry comes from the weights: in every scenario, cost conviction moves the forecast more than demand conviction (κ_C of 1.19, 0.61 and 0.86, against κ_D of 0.00, 0.47 and 0.32), so excluding respondents who tick the cost channel lowers the demand gap more than the reverse. In the oil-price scenario, the demand weight is estimated at zero, and demand conviction acts only through the product term. Figure E.2 draws the estimates, pooled over the three scenarios, in the layout of Figure 3: the forecast curve of the cost channel rises steeply already at low conviction ranks, whereas that of the demand channel stays below 30 percent up to the threshold for writing. Counting write-only respondents as writers, as the model does, the pooled raw changes are 8.3 points for the cost channel and -9.9 points for the demand channel; the model reproduces both, with changes of 6.1 and -8.6 points.

Table E.5: Structural decomposition of the format effect in Andre et al. (2022)

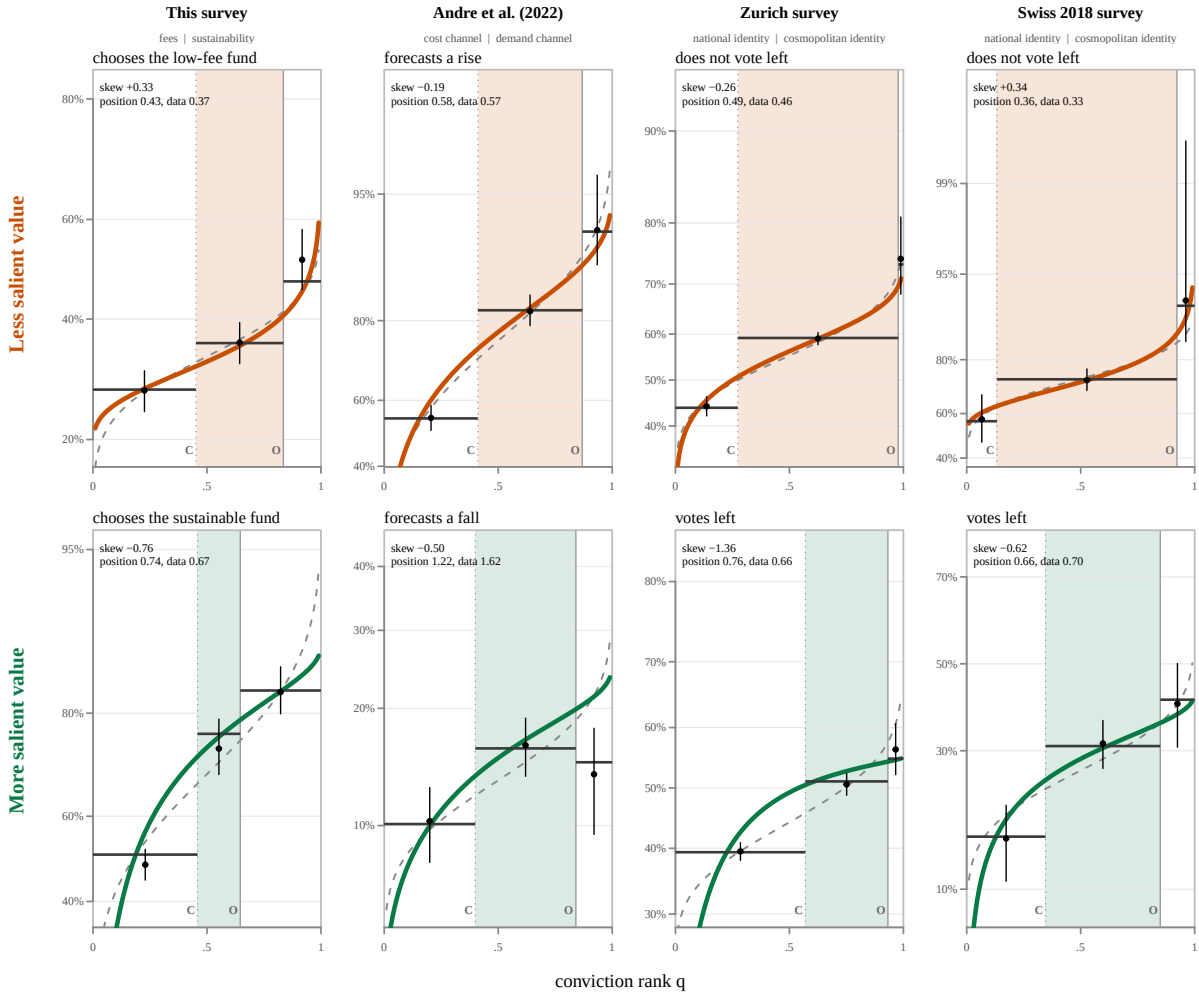
	Model (percentage points)			Raw
Scenario	Broadening	Screening	Change	change
Panel A: Cost channel as the target				
Oil price	8.1	0.8	8.9	7.8
	(1.8)	(0.4)	(1.9)	
Interest rate	6.0	−3.5	2.5	5.7
	(1.9)	(2.1)	(2.8)	
Income tax	8.3	−1.5	6.8	11.3
	(2.2)	(1.2)	(2.2)	
Pooled	7.5	−1.3	6.1	8.3
	(1.8)	(1.0)	(1.9)	
Panel B: Demand channel as the target				
Oil price	0.5	−16.6	−16.1	−21.2
	(1.4)	(3.8)	(3.1)	
Interest rate	7.2	−10.3	−3.1	−8.5
	(4.2)	(2.9)	(3.5)	
Income tax	3.5	−9.7	−6.1	0.6
	(2.3)	(2.1)	(1.9)	
Pooled	3.7	−12.3	−8.6	−9.9
	(2.2)	(2.5)	(1.8)	

Note: This table fits the model of Section 6 to wave 3 of Andre et al. (2022), with a Gaussian copula between the two channels' conviction ranks. Broadening is the change in the model-implied gap when the target channel moves from written to ticked reports, and screening the change when the other channel counts as unreported only if neither written nor ticked; each averages the two paths of these changes, as in equation (11), and the change is their sum. Raw changes are the corresponding changes in the data, counting write-only respondents as writers. The skews, −0.19 for the cost channel and −0.50 for the demand channel, and the product coefficients, −0.33 for a rising and −0.54 for a falling forecast, are shared across scenarios. In the bootstrap, a skew reaches its bound in 2 percent of the samples. Standard errors in parentheses are standard deviations over 2,000 bootstrap resamples of respondents within scenarios; pooled values are averages weighted by scenario size.

The model does not fit every cell: a likelihood-ratio test against the unrestricted cell shares of the three scenarios rejects it ($p < 0.001$). In simulated samples of the actual size drawn from the estimates, the estimator recovers the four pooled margins with biases of at most 0.4 points, whereas the skews are recovered less precisely; the comparison therefore rests on the margins, not on the skews or the weights.

Three differences limit the comparison with this paper's design. First, respondents always wrote before they ticked, so the ticked reports are second answers, and the comparison corresponds to the paired answers in Appendix B.8, not to the randomized

Figure E.2: Choice curves in the four surveys



Bold: the model; dashes: no skew; segments: the model's group average; dots: observed group rate, 95% interval. Log-odds axis, labelled in percent.

Note: The panels draw, in the layout of Figure 3, the probability of acting on the less salient value (top) and on the more salient value (bottom) at each conviction rank: in this survey's open-first group, whose respondents write before they tick, and in the surveys of Appendices E.3 and E.4.1, pooled over scenarios or countries with weights proportional to sample size. Bold curves are the model, with the same specification in all four surveys; grey dashes are the model with normal scores and without the product term. Grey segments are the model's average over the respondents who report neither, only tick, or write, drawn across each group's ranks, and black dots the observed rates of these groups, with 95% intervals. The curves integrate the other value's rank over its distribution; with correlated ranks, as in the other surveys, the group averages need not lie on the curves. Shading between C and O marks the respondents who tick the value without writing about it, at their observed share. Each panel gives the value's skew and the ticked-only group's position between non-reporters (0) and writers (1) on the log-odds axis, in the model and in the data. The vertical axis is the log odds, labelled in percent.

first-position comparison. Second, both reports explain a forecast that respondents had already made, so they may partly rationalize it. Rationalized written answers would follow the forecast more often than ticked ones, which could add to the rise of the cost channel's gap but cannot produce the fall of the demand channel's gap. Third, the outcome is a forecast rather than a choice, and the written codes and list items describe related but not identical channels.

E.4 Evidence from further surveys

The surveys are those of Table 2. For the surveys with a binary outcome, I fit the model of Section 6 with two values and two outcomes. As for the inflation forecasts, the conviction ranks are linked by a Gaussian copula, respondents who write about a value without ticking it count as writers, the outcome is a logit in the two convictions' indices and their product, as in eq. (6) without the cards, with the skews and the product coefficient shared across the countries of the Zurich survey, and I estimate the model by maximum likelihood, here on the 18 combinations of the two reporting levels and the outcome. The model has no choice cards, earliness groups, or priming.

E.4.1 Identity surveys

In the Social Identities and Cleavage Formation Survey (Bornschier et al. 2024) (Section 3.2), closeness is rated on a scale from 1 to 10. Before the open question, randomized treatment groups saw pictures that primed particular groups; I pool all groups. I use the release in the replication package of Zollinger and Häusermann (2026) and the 7,886 respondents whose party belongs to the authors' party families. A written report is a description that contains national or cosmopolitan words from multilingual keyword lists compiled for this paper; a ticked report is a closeness of at least 7 to the country's people or to cosmopolitans. The outcome is a vote for a party of the left bloc rather than the center right or the far right: cosmopolitan identity points toward the left, national identity away from it. The Swiss survey of 2018 (Bornschier et al. 2021, Zollinger 2024) asked the same open question first and closeness on a four-point scale (ticked: rather or very close), and no respondent saw a priming picture; 656 respondents name a party in the same families.

Table E.6 reports the raw comparison, counted as in Table 3. In every country, the list raises the gap of national identity one identity at a time and lowers the gap of cosmopolitan identity net of national identity. Net of cosmopolitan identity, the national gap rises in Germany and the United Kingdom but falls in Switzerland and France, so the pooled rise rests on the first two. The Swiss survey gives similar changes, with larger standard errors for cosmopolitan identity. In the control groups of the Zurich survey, which saw no pictures before the open question, all four changes keep their signs.

Table E.6: Written and ticked identities and the vote in the identity surveys

	Reports (%)		Gap, one identity (%)			Gap, net of the other (%)		
Country	Written	Ticked	Written	Ticked	Change	Written	Ticked	Change
Panel A: National identity (vote not left)								
Switzerland	3.5	74.5	19.5	31.4	11.9	20.9	20.2	−0.7
					(4.6)			(5.2)
Germany	2.3	74.1	23.7	45.5	21.8	22.2	38.3	16.1
					(7.0)			(7.0)
France	3.1	70.1	31.7	32.8	1.1	30.8	22.9	−7.8
					(7.2)			(7.4)
United Kingdom	0.9	70.6	37.5	48.6	11.1	37.5	47.1	9.6
					(10.0)			(10.0)
Pooled	2.3	72.4	28.5	40.5	12.0	28.4	33.6	5.2
					(4.1)			(4.2)
Swiss survey, 2018	7.8	86.1	7.8	25.0	17.1	6.5	14.7	8.2
					(4.0)			(4.2)
Panel B: Cosmopolitan identity (vote left)								
Switzerland	9.1	52.2	54.4	55.4	1.0	52.6	48.4	−4.2
					(3.4)			(4.2)
Germany	11.3	47.2	33.5	39.8	6.3	33.3	22.8	−10.5
					(3.4)			(4.1)
France	5.3	43.1	44.9	53.0	8.1	44.8	40.2	−4.6
					(5.9)			(7.0)
United Kingdom	2.8	25.8	38.2	41.7	3.6	38.2	26.1	−12.0
					(5.8)			(6.4)
Pooled	6.8	40.7	42.9	47.1	4.2	42.3	34.0	−8.3
					(2.5)			(2.9)
Swiss survey, 2018	15.1	63.3	59.6	66.3	6.7	58.5	50.8	−7.7
					(4.4)			(7.3)

Note: This table uses the Social Identities and Cleavage Formation Survey (Bornschier et al. 2024), in the release of Zollinger and Häusermann (2026) (7,886 voters whose party belongs to the authors' families), and the Swiss survey of 2018 (Bornschier et al. 2021) (656 voters). Written reports are national or cosmopolitan words in the description of people who are like the respondent; ticked reports are a closeness of at least 7 on the 1–10 scale (2018 survey: rather or very close). The gap is the share of an identity's reporters whose vote does not follow the identity: a vote for the left for national identity, a vote for another party for cosmopolitan identity. Net of the other identity, it uses reporters who do not report the other identity in the same format. Changes are ticked minus written, for the same respondents. Pooled values are averages over the four countries weighted by sample size. Standard errors in parentheses are standard deviations over 2,000 bootstrap resamples of respondents within countries.

Table E.7 reports the model's decomposition. For national identity, broadening dominates (11.2 against −5.2 points), and the model's pooled change, 5.9 points, is close to the raw change counted as in the model, 5.2 points. For cosmopolitan identity, the stricter screen of national identity dominates (−12.0 against 2.1 points), and the model's pooled change, −9.9 points, is close to the raw change, −8.1 points. In the Swiss survey, the model's changes also come close to the raw ones (6.5 against 7.2 points for national identity and −5.9 against −5.4 for cosmopolitan identity). In the Swiss survey, the model fits the cell shares (likelihood-ratio test against the unrestricted shares: $p = 0.27$); in the four-country survey, the test rejects it ($p < 0.001$).

Figure E.2 draws these estimates, pooled over the four countries of the Zurich survey, in the layout of Figure 3. The shading between C and O marks the respondents whom the list adds to an identity's reporters. The more a curve rises across this interval, the more these respondents vote differently from the writers, and the larger the broadening. Across countries, it is largest for national identity in Germany and the United Kingdom and for cosmopolitan identity in Switzerland and France (Table E.7).

The written reports rest on the keyword lists. I checked them against my own coding of 500 answers from the Zurich survey, which I coded by meaning with a written codebook, without seeing the closeness ratings, the vote or the lists' matches. A written identity names the respondent's own nation or a cosmopolitan outlook, such as openness to the world or to other cultures, or tolerance. Traits that tend to go with an identity, such as being down-to-earth, open-minded or fond of travel, do not count, and neither does a party. The lists find about two thirds of the identities I coded, and about two thirds of their matches are identities by my coding (Table E.8; Cohen's κ of 0.67 for national and 0.60 for cosmopolitan identity). Their errors largely cancel in the shares of writers: 2.3 and 6.5 percent with the lists, 2.6 and 6.4 percent in my coding. On the coded answers, broadening again dominates for national identity and the screen for cosmopolitan identity, and no statistic differs significantly from its value with the lists.

The lists also count words for open-mindedness and travel, which my coding does not. Without these words, the national changes and margins move by at most 0.4 points and the cosmopolitan ones by up to 4.3 points. All keep their signs, and the screen of cosmopolitan identity still exceeds its broadening; net of national identity, its change is −4.7 instead of −8.3 points in the Zurich survey and −6.3 instead of −7.7 in the Swiss survey.

Because both surveys rate closeness to the two groups on one scale, they show whether writing about a group takes the same closeness for both identities. It does not (Figure E.3). More respondents write about cosmopolitan identity than about national identity at every level of closeness but the lowest of the Swiss survey, where very few

Table E.7: Structural decomposition of the format effect in the identity surveys

	Model (percentage points)			Raw
Country	Broadening	Screening	Change	change
Panel A: National identity as the target				
Switzerland	6.6	−8.1	−1.5	−0.9
	(2.2)	(1.1)	(2.9)	
Germany	16.7	−4.9	11.9	16.3
	(4.5)	(1.2)	(4.9)	
France	4.0	−8.8	−4.9	−8.0
	(2.3)	(1.3)	(3.1)	
United Kingdom	15.0	−1.3	13.6	9.7
	(5.4)	(0.9)	(5.6)	
Pooled	11.2	−5.2	5.9	5.2
	(3.5)	(0.8)	(3.8)	
Swiss survey, 2018	12.1	−5.6	6.5	7.2
	(3.8)	(2.2)	(4.6)	
Panel B: Cosmopolitan identity as the target				
Switzerland	3.6	−10.3	−6.7	−4.9
	(2.5)	(2.0)	(3.0)	
Germany	2.4	−14.2	−11.9	−9.6
	(1.2)	(2.0)	(2.4)	
France	3.3	−8.9	−5.6	−3.8
	(3.2)	(2.8)	(4.5)	
United Kingdom	0.2	−13.7	−13.5	−11.8
	(0.7)	(2.2)	(2.5)	
Pooled	2.1	−12.0	−9.9	−8.1
	(1.7)	(1.4)	(2.3)	
Swiss survey, 2018	7.1	−13.0	−5.9	−5.4
	(4.2)	(5.7)	(6.2)	

Note: This table fits the binary model of Appendix E.4, with a Gaussian copula between the two identities' conviction ranks, to the surveys of Table E.6. Broadening is the change in the model-implied gap when the target identity moves from written to ticked reports, and screening the change when the other identity counts as unreported only if neither written nor ticked; each averages the two paths of these changes, as in equation (11), and the change is their sum. Raw changes are the corresponding changes in the data, counting write-only respondents as writers. The skews of national and cosmopolitan identity and the product coefficient are −0.26, −1.36 and −0.03 in the Zurich survey, shared across countries, and 0.34, −0.62 and 0.00 in the Swiss survey. In the bootstrap, a skew reaches its bound in 30 percent of the samples for the Zurich survey and 23 percent for the Swiss survey. Pooled values are averages over the countries weighted by sample size. Standard errors in parentheses are standard deviations over 2,000 bootstrap resamples of respondents within countries.

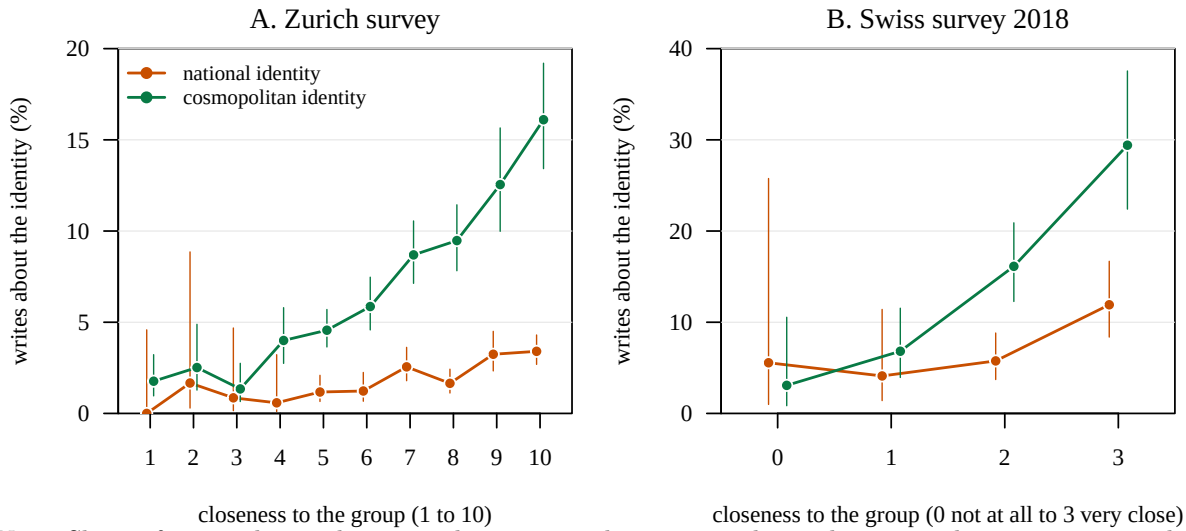
Table E.8: The identity keyword lists against hand codes

Panel A: Agreement of the lists with the hand codes					
	Hand-coded	Precision	Recall	F_1	κ
National identity	28	72.0	64.3	67.9	0.67
		(9.1)	(9.2)	(7.6)	(0.08)
Cosmopolitan identity	63	62.0	62.8	62.4	0.60
		(5.8)	(7.1)	(5.2)	(0.06)
Panel B: Statistics on the coded answers (points)					
	Lists	Hand codes	Difference		
Writers, national identity (%)	2.3	2.6	0.3		
	(0.4)	(0.5)	(0.4)		
Writers, cosmopolitan identity (%)	6.5	6.4	−0.1		
	(0.6)	(0.8)	(0.8)		
Writing difference at closeness 9–10	18.2	15.3	−2.8		
	(4.1)	(4.0)	(2.8)		
National identity					
Gap change, one identity	10.8	20.9	10.1		
	(9.4)	(7.6)	(6.9)		
Gap change, net of the other	13.2	23.8	10.6		
	(10.1)	(8.3)	(7.2)		
Broadening	12.6	19.9	7.3		
	(10.1)	(8.5)	(7.0)		
Screening	1.0	3.3	2.3		
	(4.3)	(2.8)	(3.1)		
Cosmopolitan identity					
Gap change, one identity	10.1	16.7	6.5		
	(6.2)	(6.9)	(5.6)		
Gap change, net of the other	−26.0	−19.4	6.6		
	(7.5)	(8.0)	(5.6)		
Broadening	−0.6	2.3	2.9		
	(6.5)	(6.7)	(4.9)		
Screening	−22.6	−19.7	2.9		
	(5.5)	(5.6)	(3.6)		

Note: The table compares the keyword lists with my hand codes of 500 descriptions of people who are like the respondent in the Zurich survey. The 500 are drawn from a random sample of 1,200 of the respondents of Table E.6: 250 among the 276 answers that the lists, broader keyword lists or a coding by a language model flag, and 250 among the other 924. Each answer is weighted by the size of its stratum, so that the statistics refer to the sample of 1,200. The hand codes come from one coder, blind to closeness, vote and the lists' matches, with a second reading of all 500 against the codebook. Panel A: hand-coded are the answers coded as naming the identity; precision is the share of the lists' matches that the hand codes confirm, recall the share of the hand-coded identities that the lists find, and F_1 their harmonic mean, all in percent; κ is Cohen's κ. Panel B computes the statistics of Tables E.6 and E.9 on the coded answers, pooled over countries, with the lists and with the hand codes; the writing difference at closeness 9–10 is the share writing about cosmopolitan identity among respondents with closeness 9 or 10 to cosmopolitans minus the same share for national identity. Standard errors in parentheses are standard deviations over 2,000 bootstrap resamples within the two strata.

write at all, and the difference grows with closeness, to 16 against 3 percent at the top of the Zurich scale and 29 against 12 percent among the very close in the Swiss survey. The closed-ended threshold is the same for both identities, so the writing threshold lies further above it for national identity, the less salient one (Section 6.3). The pattern holds in the control groups of the Zurich survey and in my coding: among the coded answers with closeness 9 or 10, the difference is 15 points, against 18 with the lists.

Figure E.3: Writing about an identity by closeness to the group



Note: Share of respondents who write about national or cosmopolitan identity in the open question, by their closeness to the country's people or to cosmopolitans, which they rate afterwards on the same scale for both groups (closed-ended report: 7 or more in the Zurich survey, rather or very close in the Swiss survey), with 95% intervals. Written identities are found with the keyword lists. Samples as in Table E.6.

E.4.2 Margins measured without the model

When every respondent answers both formats, the two margins can be measured without the model, in the model's own counting. Broadening is the change in the target value's gap when its ticked-only respondents join its writers, and screening the change when respondents who tick the competing value no longer count as not reporting it; as for the model, each averages the two paths of these changes (equation (11)). Table E.9 compares these measured margins with the model's estimates. For this survey, it shows the randomized comparison of first answers, measured along the path of eq. (12), and the paired answers of the open-first group, whose respondents write before they tick, as in the other surveys; for that group, the model is evaluated at its reporting thresholds, which it fits (Table B.7). The model comes within 0.8 points of the measured broadening, except for fees in this survey's randomized comparison, the cost channel in Andre et al. (2022) and cosmopolitan identity in the Zurich survey, which it understates by 3.1, 3.7 and 2.7 points, and within 1.6 points of the measured screens. As the model assumes, respondents who only tick a value act on it more often than non-reporters and less often

Table E.9: Broadening and screening measured without the model and estimated by the model

	Broadening (points)		Screening (points)	
Value	Measured	Model	Measured	Model
This survey, randomized comparison				
Fees	12.5	9.4	−3.7	−2.6
	(3.3)	(3.8)	(1.8)	(0.6)
Sustainability	0.7	0.7	−3.6	−2.4
	(1.6)	(0.3)	(1.6)	(1.3)
This survey, open-first group				
Fees	10.1	9.3	−8.3	−7.0
	(2.9)	(3.8)	(1.6)	(1.1)
Sustainability	2.4	1.8	−3.5	−2.5
	(1.1)	(0.7)	(1.3)	(1.3)
Andre et al. (2022)				
Cost channel	11.2	7.5	−2.9	−1.3
	(2.7)	(1.8)	(1.5)	(1.0)
Oil price	9.5	8.1	−1.7	0.8
	(2.9)	(1.8)	(2.0)	(0.4)
Interest rate	13.2	6.0	−7.5	−3.5
	(3.6)	(1.9)	(3.4)	(2.1)
Income tax	11.1	8.3	0.2	−1.5
	(6.6)	(2.2)	(2.5)	(1.2)
Demand channel	3.9	3.7	−13.8	−12.3
	(4.0)	(2.2)	(2.8)	(2.5)
Oil price	−8.7	0.5	−12.5	−16.6
	(6.6)	(1.4)	(5.7)	(3.8)
Interest rate	12.4	7.2	−21.0	−10.3
	(8.7)	(4.2)	(5.4)	(2.9)
Income tax	8.8	3.5	−8.2	−9.7
	(4.7)	(2.3)	(2.8)	(2.1)
Zurich survey				
National identity	10.9	11.2	−5.7	−5.2
	(4.2)	(3.5)	(1.3)	(0.8)
Switzerland	11.8	6.6	−12.7	−8.1
	(4.0)	(2.2)	(2.7)	(1.1)
Germany	21.6	16.7	−5.3	−4.9
	(6.6)	(4.5)	(2.3)	(1.2)
France	1.2	4.0	−9.3	−8.8
	(6.9)	(2.3)	(2.6)	(1.3)
United Kingdom	8.1	15.0	1.6	−1.3
	(10.6)	(5.4)	(2.6)	(0.9)
Cosmopolitan identity	4.8	2.1	−12.9	−12.0
	(2.7)	(1.7)	(2.1)	(1.4)
Switzerland	5.5	3.6	−10.4	−10.3
	(4.0)	(2.5)	(3.3)	(2.0)
Germany	3.5	2.4	−13.1	−14.2
	(3.2)	(1.2)	(2.9)	(2.0)
France	10.3	3.3	−14.1	−8.9
	(6.3)	(3.2)	(5.0)	(2.8)
United Kingdom	2.4	0.2	−14.2	−13.7
	(6.0)	(0.7)	(4.8)	(2.2)
Swiss survey, 2018				
National identity	11.7	12.1	−4.5	−5.6
	(3.8)	(3.8)	(1.9)	(2.2)
Cosmopolitan identity	6.6	7.1	−12.0	−13.0
	(6.9)	(4.2)	(7.9)	(5.7)

Note: Measured margins use the paired answers, except this survey's randomized comparison, which combines the two question orders along the path of equation (12) (Appendix B.8). Model margins: Tables E.5 and E.7; for this survey, Section 7.2, at first-position thresholds or those of the open-first group. All margins count respondents who write about a value without ticking it as writers. The rows for each channel and identity average the indented scenarios or countries by sample size. Standard errors in parentheses: standard deviations over 2,000 bootstrap resamples within scenarios, countries or, for this survey, question order and card.

than writers when the competing value is not reported. This holds for every value in the table except the demand channel in the oil-price scenario of Andre et al. (2022), which has 12 writers.

The broadening of the less salient value falls short in this survey and in the survey of Andre et al. (2022): by 3.1 points in this survey's randomized comparison and 0.7 in its open-first group, and by 3.7 points in Andre et al. (2022). In this survey, the model overpredicts the gap of one group: in the open-first group, 31 percent of the fee writers who report nothing about sustainability choose the sustainable fund on the rating cards, and the model predicts 38 percent. With one choice per respondent, the data hardly distinguish a weak value that stays out of the choice from one that enters with a low weight (for repeated choices, see Hess et al. 2013). In the survey of Andre et al. (2022), the shortfall comes mostly from the interest-rate scenario, whose salience order reverses (Appendix E.3). For the more salient value, the negative skew, which flattens the choice index among the strongest convictions, brings the broadening close to its measured size, except for cosmopolitan identity in the Zurich survey.

The Zurich survey also allows an out-of-sample check across countries. I fit the model, with independent ranks, jointly to three of the four countries at a time, with common weights, skews and product coefficient and a country-specific intercept. For the fourth country, the reporting thresholds come from its reporting shares, and the intercept is fitted to the vote of its respondents who report neither identity; the model then predicts broadening and screening for both identities. The predictions miss the 16 measured margins by 5.4 points on average, against 3.0 points for the model fitted to all four countries and 4.8 points for the sample-weighted average of the margins measured in the other three countries. The model does better than this benchmark in the United Kingdom and worse in the other three countries; the difference, 0.6 points, has a 95 percent interval of $[-1.8, 1.3]$. With normal scores, the predictions miss the measured margins by 4.9 points, and with the hurdle version of Appendix B.5 by 3.2 points. The signs of the predictions carry little information, because with independent ranks broadening is positive and screening negative whenever the probability of a value-consistent choice rises with each conviction, as it does here; the measured margins have these signs in 15 of the 16 cases, the exception being the screen of national identity in the United Kingdom (1.6 points).

E.4.3 Prediction across question orders in this survey

The binary model also allows an out-of-sample check in this paper's randomized design. I fit it, with independent ranks and the sustainable and the conventional fund as the two outcomes, to the paired answers of one question-order group only and predict the first-position gaps of the other group. The prediction keeps the fitted weights, skews, product

coefficient and intercept, sets the reporting thresholds to the other group's reporting shares, and averages the predicted choice probabilities over exactly the respondents whose gap it predicts, given their observed reporting levels. As in the model, closed-ended reports include respondents who write about a topic without ticking it (Appendix B.8). The benchmark assumes no format effect: it predicts each gap by the fitting group's own first-position gap. Table E.10 reports the results. Fitted to the open-first group, the model predicts the closed-first gaps within 3.4 points, whereas the benchmark misses the fee gaps by 8.9 and 14.4 points. Fitted to the closed-first group, the model again predicts the sustainability gaps closely, but it overstates the open fee gaps, by 6.6 and 8.2 points, as the main estimator does at the open fee endpoint. With normal scores, the model predicts the open-first gaps slightly better (mean absolute error 3.5 against 3.9 points), but the difference is within sampling error (95 percent interval $[-1.7, 2.7]$); fitted to the closed-first group, the skews are weakly identified and reach their bound in 19 percent of the bootstrap samples.

Table E.10: Out-of-sample prediction across the randomized question orders

			No format	Absolute error	
Gap (%)	Observed	Predicted	effect	Model	No format effect
Panel A: Closed-first arm predicted from the open-first arm					
Sustainability, one value	18.0	19.2	17.0	1.1	1.1
		(1.4)			
Sustainability, net of fees	12.8	14.0	15.7	1.2	2.9
		(1.8)			
Fees, one value	62.5	60.2	48.1	2.3	14.4
		(1.7)			
Fees, net of sustainability	48.6	45.2	39.7	3.4	8.9
		(2.3)			
Panel B: Open-first arm predicted from the closed-first arm					
Sustainability, one value	17.0	16.6	18.0	0.4	1.1
		(1.6)			
Sustainability, net of fees	15.7	15.5	12.8	0.2	2.9
		(1.6)			
Fees, one value	48.1	54.7	62.5	6.6	14.4
		(3.4)			
Fees, net of sustainability	39.7	47.9	48.6	8.2	8.9
		(3.8)			

Note: The binary model of Appendix E.4, with independent ranks and without card, earliness, or priming components, is fitted to the paired answers of one question-order group and predicts the first-position gaps of the other group. Closed-ended reports include respondents who write about a topic without ticking it, as in the model. The no-format-effect benchmark predicts each gap by the fitting group's own first-position gap. Standard errors in parentheses are standard deviations of the prediction over 2,000 bootstrap resamples of respondents within question order and card, refitting the model each time.

F Robustness Checks Tables

Like Schuman and Presser (1979), I analyzed the relation of socioeconomic background variables with the sustainability and fees topics in the closed and open-ended responses.[22] Table F.1 shows the results of an OLS regression with the topic dummies for the closed and open-ended topics as dependent variables (in each regression, the sample is restricted to respondents answering the question type first).

For the sustainability topic, both response types show a significant and similar positive coefficient for women, higher education, and altruism. There are differences between the two question types' association with background variables. While the general importance of mitigating climate change and having made a donation is positive for both question types, the magnitude is smaller for the open-ended response. For the fees topic, both question types have a significant negative coefficient for women, age, and altruism. Financial literacy is positive and significant for both question types. However, the magnitude is three times as large for closed-ended responses; only closed-ended responses show a positive coefficient for time preferences. In contrast, only the open-ended fees topic shows a positive and significant coefficient for income and wealth.

As a complementary check, I regressed sustainable-fund choice on the socioeconomic variables instead of the reported topics (Table F.1, column 5). The coefficients match the ones that explain the two relevant topics but more strongly those from the open-ended responses. Age, female gender, donation, altruism, and the importance of climate change are statistically significant and positive, while financial literacy, income, wealth, and risk preferences are statistically significant and negative.

As a final robustness check, I show, using a multinomial logistic regression, the correlation between the background variables and the three types of closed-ended responses: equivalent, non-equivalent, and, as the reference group, no closed-ended topic. Table F.2 presents the results. The results indicate that for many background characteristics, the effect is similar for both equivalent and non-equivalent answers. However, the magnitude of many coefficients is larger for equivalent responses. Some variables, such as donating, are only statistically significant for the corresponding sustainability topic. For fees, only the coefficients for the equivalent topic are statistically significant for gender, pensioner status, wealth, risk, and trust. This pattern indicates a stronger association for equivalent topics with the background variables and a greater distance from the reference group, which consists of respondents who did not choose any closed-ended topic.

[22] Appendix Tables G.1 and G.2 define the socioeconomic variables.

Table F.1: OLS Regressions with demographics

	Sust closed	Sust open	Fees closed	Fees open	sust invest
Financial literacy	0.011	0.032	0.108***	0.030*	−0.026*
	(0.023)	(0.022)	(0.025)	(0.017)	(0.015)
Age	0.001	0.003**	−0.003***	−0.004***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female	0.102***	0.091***	−0.072**	−0.074***	0.034*
	(0.031)	(0.030)	(0.033)	(0.024)	(0.021)
University educ	0.106***	0.089***	0.006	−0.014	−0.020
	(0.030)	(0.030)	(0.032)	(0.024)	(0.021)
Log income	−0.037	0.028	0.027	0.054*	−0.049*
	(0.038)	(0.037)	(0.040)	(0.029)	(0.025)
Log wealth	−0.014	−0.019	0.013	0.039***	−0.032***
	(0.015)	(0.014)	(0.016)	(0.011)	(0.010)
Donation	0.097***	0.056	−0.017	0.019	0.106***
	(0.037)	(0.036)	(0.039)	(0.029)	(0.025)
Risk preference	−0.007	0.006	0.012*	0.003	−0.011**
	(0.007)	(0.006)	(0.007)	(0.005)	(0.004)
Time preference	0.004	−0.004	0.023***	0.009	0.003
	(0.008)	(0.008)	(0.008)	(0.007)	(0.006)
Altruism	0.024***	0.034***	−0.019**	−0.013**	0.015***
	(0.007)	(0.008)	(0.007)	(0.006)	(0.005)
Trust	0.001	0.001	0.012*	−0.004	0.005
	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)
Importance climate	0.054***	0.032***	−0.006	0.005	0.046***
	(0.006)	(0.007)	(0.007)	(0.005)	(0.004)
Num.Obs.	1036	1123	1036	1123	2159
R2	0.168	0.111	0.074	0.067	0.124

Note: This table shows OLS regressions of closed- and open-ended sustainability mentions (columns 1–2), closed- and open-ended fee mentions (columns 3–4), and sustainable-fund choice (column 5) on respondent characteristics. Columns (1)–(4) use respondents who answered that format first, column (5) all respondents; respondents who did not report income or wealth are excluded. Classical standard errors in parentheses; no fixed effects. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table F.2: Multinomial logistic regressions, closed-ended sustainability and fee topics

	Sustainability topic (closed)		Fees topic (closed)	
	equivalent	non-equivalent	equivalent	non-equivalent
Financial literacy	0.136	−0.023	1.209***	0.259**
	(0.138)	(0.157)	(0.223)	(0.116)
Age	0.005	−0.002	−0.016*	−0.009
	(0.008)	(0.009)	(0.008)	(0.007)
Female	0.498***	0.404*	−0.581***	−0.206
	(0.175)	(0.208)	(0.194)	(0.159)
University educ	0.473***	0.493**	0.005	0.022
	(0.171)	(0.206)	(0.183)	(0.156)
Pensioner	0.184	0.276	−0.564*	0.104
	(0.274)	(0.339)	(0.323)	(0.257)
Married	−0.097	−0.092	−0.039	−0.089
	(0.200)	(0.240)	(0.219)	(0.181)
HH-size	−0.003	−0.045	−0.120	0.010
	(0.079)	(0.097)	(0.087)	(0.073)
Log income	−0.159	0.206	0.177	0.123
	(0.228)	(0.272)	(0.251)	(0.204)
Log wealth	−0.106	−0.039	0.217**	−0.044
	(0.085)	(0.102)	(0.089)	(0.080)
Donation	0.924***	0.085	0.100	−0.153
	(0.247)	(0.243)	(0.228)	(0.189)
Risk preference	−0.029	−0.067	0.095**	0.025
	(0.038)	(0.046)	(0.042)	(0.035)
Time preference	0.059	−0.054	0.131***	0.083**
	(0.046)	(0.052)	(0.049)	(0.040)
Altruism	0.170***	0.077	−0.112**	−0.065*
	(0.042)	(0.048)	(0.044)	(0.037)
Trust	−0.009	0.083**	0.078**	0.037
	(0.032)	(0.039)	(0.035)	(0.030)
Climate Awareness	0.424***	0.248***	−0.026	−0.022
	(0.057)	(0.054)	(0.039)	(0.033)
Num.Obs.	1036		1036	
R2 (McFadden)	0.380		0.348	

Note: Coefficients are multinomial logistic-regression coefficients (log-odds), estimated separately for the closed-ended sustainability topic (columns 1–2) and the closed-ended fees topic (columns 3–4). Categories: equivalent closed-ended mention, non-equivalent closed-ended mention; the reference category is no closed-ended mention. Standard errors in parentheses. McFadden pseudo-R^2. Sample: respondents answering the closed-ended question first, with non-missing income and wealth ($N = 1{,}036$; smaller than the full $N = 3{,}059$ for that reason). $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

G Background Variables

Table G.1: Definition of variables obtained from the survey company

Variable	Description
Female	A dummy variable that takes the value of 1 if the respondent identifies as female and 0 otherwise.
Age	Age in years.
University Degree	A dummy variable that takes the value of 1 if the respondent holds a university degree and 0 otherwise.
Pensioner	A dummy variable that takes the value of 1 if the respondent's current employment status is "pensioner" and 0 otherwise.
Married	A dummy variable that takes the value of 1 if the respondent is married and 0 otherwise.
Household size	The number of people currently living in the respondent's household.
Income	Respondents were classified into six intervals for their monthly income: "less than 3,000 CHF," "3,000–4,500 CHF," "4,501–6,000 CHF," "6,001–9,000 CHF," "9,001–12,000 CHF," and "more than 12,000 CHF." I converted the intervals into a continuous variable by taking the average value for the interval when possible. The new continuous income variable thus takes the values: 3,000 CHF, 3,750 CHF, 5,250 CHF, 7,500 CHF, 10,500 CHF, and 14,000 CHF.
Wealth	Respondents were classified into seven intervals for their wealth: "less than 50,000 CHF," "50,000–100,000 CHF," "100,000–250,000 CHF," "250,000–500,000 CHF," "500,000–1,000,000 CHF," "1,000,000–5,000,000 CHF," and "more than 5,000,000 CHF." I converted the intervals into a continuous variable by taking the average value for the interval when possible. The new continuous wealth variable thus takes the values: 50,000 CHF, 75,000 CHF, 175,000 CHF, 375,000 CHF, 750,000 CHF, 3,000,000 CHF, and 5,000,000 CHF.

Note: This table describes the variables the survey company collected on its panel members. The survey company provided these background variables on the sample of respondents used in this paper. All tables use the values from the 2021 wave, except Table D.3, which uses the same variables from a 2024 update of the panel.

Table G.2: Definition of variables from the survey

Variable	Description
Financial Literacy	The "big three" questions on financial literacy by Lusardi and Mitchell (2008).
Risk Preferences	"How willing or unwilling are you to take risks when making decisions in your life?" (0 = "Completely unwilling to take risks," 10 = "Very willing to take risks"), following Falk et al. (2023).
Time Preferences	"How willing are you to give up something that is beneficial for you today to benefit more from that in the future?," on a 10-point scale (1 = "Completely unwilling" to 10 = "Very willing to do so"), following Falk et al. (2023).
Altruism	"How would you assess your willingness to share something with others without expecting anything direct and immediate in return?" (1 ="Completely unwilling" to 10 = "Very willing to do so"), following Falk et al. (2023).
Trust	"As long as I am not convinced otherwise, I always assume that other people only have the best in mind." (0="No approval at all," 10="Very strong approval"), following Falk et al. (2023).
Climate Awareness	"Climate change is a serious problem that needs to be solved." (0 = "No approval at all," 10 = "Very strong approval"), following Heeb et al. (2023).
Donate	"Yes" to either: "Have you made at least one donation to a social institution in the last 12 months (e.g., Salvation Army, Swiss Solidarity, SOS Children's Villages, etc.)?" or "Have you made at least one donation to an environmental organization in the last 12 months (e.g., Greenpeace, WWF, myclimate, etc.)?"

Note: This table describes the variables collected directly in the survey. All variables were collected after the respondents answered the open and closed questions.