Green Tastes, Beliefs, and Portfolios: Evidence from Synthetic Investors

Xindi He Lubos Pastor*

October 26, 2026

Abstract

We use LLMs as synthetic survey respondents to elicit green tastes, return beliefs, portfolio allocations, and investment motives from 480 retail and 60 institutional investor identities. Compared with retail, institutions report weaker tastes yet greener portfolio tilts, citing outside pressure. A greener tilt therefore need not indicate a stronger taste for greenness. Retail tastes depend mostly on political affiliation, institutional tastes on type and geography. Beliefs differ across the three ESG dimensions, with high-governance portfolios expected to outperform. We reconstruct survey responses under 2014–2025 information environments using date-masked information modules shuffled across years. Tastes vary little over time, but beliefs about green portfolios become more pessimistic after 2020. The two reconstructed series relate to different real-world data: tastes to measures of green demand, beliefs to measures of attention.

JEL classifications: G11, G23, G41, Q56, C83

Keywords: large language models, synthetic surveys, sustainable investing, non-pecuniary preferences, investor beliefs, portfolio choice, institutional investors

---

*He is at the Georgia Institute of Technology. Pastor is at the University of Chicago Booth School of Business, NBER, and CEPR. Pastor also serves as an independent director and trustee of Vanguard. The views expressed here do not necessarily reflect those of Vanguard or its funds. We thank Alessandro Dodon for excellent research assistance.

1 Introduction

Investors' non-pecuniary preferences, or tastes, for green assets are a key input to theories of sustainable investing (e.g., Pastor et al., 2021; Pedersen et al., 2021). They are difficult to measure, though, because observed green portfolio tilts need not reflect tastes alone—investors may also tilt green because they expect green portfolios to outperform, or because their mandates require it. We provide survey evidence that non-taste drivers matter: institutional respondents state weaker green tastes than retail respondents, yet they allocate more to the greener portfolio in a hypothetical allocation task. Tastes therefore cannot be inferred from green portfolio tilts; they must be elicited separately from beliefs and mandates.

We use large language models (LLMs) as synthetic survey respondents to elicit four objects separately from the same investor identities: green tastes, return beliefs, desired portfolio allocations, and the motives behind a green holding. We present respondents with two hypothetical equity portfolios, H and L, that are matched on all characteristics except for their sustainability scores. To measure tastes, we ask respondents to state the return they would give up to hold the high-score Portfolio H and the additional return they would require to hold the low-score Portfolio L; we compute willingness to pay for greenness (WTP) as the average of the two. To measure beliefs, we elicit the respondents' expectations of returns for these matched portfolios as well as unmatched typical real-world portfolios with high and low sustainability scores. We also elicit the respondents' desired allocations to the two portfolios and ask whether a green holding reflects the respondent's own choice or outside pressure. Each module contains two questions and is administered in a separate model call, without questions or answers from other modules. This design separates the three potential sources of a green tilt within the same identity: tastes, beliefs, and outside pressure.

Our survey covers 480 retail identities and 60 institutional identities. The retail identities form a complete grid over political affiliation, income, age, gender, and financial literacy, allowing each characteristic to vary independently of the others. The institutional identities form an analogous grid over institution type, size, and geography. Each institutional identity answers as the investment decision-maker at that institution. We vary the sustainability dimension one at a time, allowing the responses to environmental, social, and governance (ESG) scores to be compared under the same design. We separately vary the economic state to study how tastes and beliefs change across economic conditions. Four models, including

our main model, ChatGPT-5.6 Sol, complete the full survey, and nine additional model versions answer common subsamples.

Looking across investor types, we find that greener portfolio allocations do not imply stronger green tastes. In the main model, institutional WTP is 0.18% per year, compared with 0.49% for retail identities. Institutions are also, if anything, more pessimistic than retail identities about the returns on green portfolios. Yet institutions allocate 70% of their hypothetical equity portfolios to Portfolio H, compared with 60% for retail identities. These comparisons have the same sign in all four models. In other words, compared with retail, institutions state weaker green tastes yet allocate more to the green portfolio.

Stated motives help explain this apparent puzzle. Retail identities attribute more than 80% of their Portfolio H holding to their own choice in every model, whereas institutions attribute roughly half to outside pressure. Reaching this conclusion requires eliciting tastes, beliefs, allocations, and motives from each identity, with the same questions for retail and institutional investors—a combination difficult to achieve in human surveys.

Looking within investor types, we find that green tastes are closely related to investor characteristics. Among retail identities, political affiliation is by far the strongest predictor of WTP. In the main model, average WTP falls from 1.04% per year for strong Democrats to 0.02% for strong Republicans, and political affiliation ranks first in all 13 models. Income and financial literacy have smaller, opposing effects: holding other characteristics fixed, WTP rises with income but falls with literacy. This pattern can be obscured in human samples, where income and literacy are positively correlated (Lusardi and Mitchell, 2014). Among institutional identities, type and geography matter most: endowments and foundations state higher WTP than hedge funds and mutual fund companies, and European institutions state higher WTP than U.S. institutions. Economic conditions also affect the two investor types differently: a recession lowers WTP for retail but not for institutions.

Investors differ also in their beliefs about green portfolios. We elicit two belief measures. Both measure the expected return on a high-score portfolio minus that on a low-score portfolio. The *matched belief gap* compares Portfolios H and L, which differ only in sustainability score, and thus captures beliefs about greenness itself. The *unrestricted belief gap* instead compares typical real-world high- and low-score portfolios, which may also differ in industry composition, valuations, and other characteristics. Among retail identities, the two gaps have similar near-zero means, but the unrestricted gap has a larger cross-identity standard deviation

(0.90% versus 0.49%). Removing the matching restriction therefore widens disagreement across identities without materially changing the average belief gap.

Tastes and beliefs respond differently across the three ESG dimensions. In all four models, respondents on average expect high-governance portfolios to outperform otherwise identical low-governance portfolios (Gompers et al., 2003). Relative to aggregate ESG, a governance-only score raises the matched belief gap by 0.39 to 1.09 percentage points. The environmental and social dimensions have smaller effects on beliefs. For tastes there is no consistent pattern. Environmental and social scores change WTP little, and a governance-only score raises WTP in the two Claude models and lowers it in the two ChatGPT models.

We also study variation over time by holding the main model and seven identities fixed while varying the information environment. Retrospective elicitation creates a look-ahead concern because the model has been trained on information from later periods. We address this concern by dividing each year's information environment into modules, masking identifying temporal cues, and mixing modules across years. We use the estimated module effects to reconstruct tastes and beliefs for each year and relate the resulting series to real-world measures of sustainable investing and of attention to climate and ESG.

Across the reconstructed 2014–2025 information environments, we find that tastes vary little, whereas beliefs about typical real-world green portfolios become more pessimistic after 2020. From 2020 to 2023, the expected return on such portfolios relative to brown ones falls by about 25 basis points (bps), while reconstructed WTP stays within a range of four bps. Tastes move most closely with the policy-and-regulation module, while beliefs move most closely with the economy-and-markets module. Both patterns are stronger than estimation noise alone would generate.

The two reconstructed series relate to different real-world data. Tastes move with quantities and prices in sustainable investing: net flows into U.S. sustainable funds, subsequent fund launches, and the fee that investors are willing to pay for ESG index funds. Beliefs move inversely with attention to climate and ESG in the media, in Google searches, and on earnings calls. This direction is consistent with higher green valuations and lower expected returns when attention raises demand for green assets.

Our use of LLMs as synthetic respondents builds on the synthetic-survey literature (Argyle et al., 2023; Horton et al., 2026; Brand et al., 2026; Fedyk et al., 2026; Bhagwat et al., 2026). Section 2.2 compares our design with the closest studies.

Our first contribution is to compare retail and institutional investors within the same survey design. Existing surveys in sustainable investing typically examine the two groups separately (Riedl and Smeets, 2017; Bauer et al., 2021; Heeb et al., 2023; Giglio et al., 2025; Krueger et al., 2020; Edmans et al., 2026). To our knowledge, no other academic study poses the same questions about green tastes and return beliefs to retail investors and institutional decision-makers.[1] This comparison yields our main result: the greener institutional tilt comes with weaker green tastes and more outside pressure than the retail tilt.[2]

Our second contribution is to measure green tastes separately from return beliefs and to relate both to portfolio choice. The tastes that theories of sustainable investing take as given (Pastor et al., 2021, 2025b) are difficult to identify from market outcomes: holdings and fund flows can reflect return expectations or hedging demand as well as tastes, while prices and returns reflect their equilibrium effects (Starks, 2023; Pedersen et al., 2021; Manski, 2004; Engle et al., 2020; Hartzmark and Sussman, 2019; Pastor et al., 2022, 2025a; Baker et al., 2024; Eskildsen et al., 2026). Experiments address this problem by eliciting WTP while holding financial beliefs fixed by design (Heeb et al., 2023; Bonnefon et al., 2025; Brodback et al., 2026). A large survey literature directly elicits investors' return expectations and relates them to portfolios and fund flows (Greenwood and Shleifer, 2014; Giglio et al., 2021; Dahlquist and Ibert, 2024). Surveys of individual investors also relate sustainable-fund holdings to both return expectations and non-pecuniary motives (Riedl and Smeets, 2017; Giglio et al., 2025). We bring these approaches together by eliciting tastes and beliefs separately and directly, both in units of annual return, from the same investor identities.

Separating tastes from beliefs shows that the two play different roles. Green portfolio tilts are more closely tied to tastes than to beliefs about greenness itself: when we regress the Portfolio H share on both WTP and the matched belief gap, the WTP coefficient is larger in all four models and about twice as large in the main model. Tastes and beliefs also respond differently to governance scores and to the information environment after 2020, and they

[1] The industry survey of CFA Institute (2022) asks both groups common questions about their interest in and motives for ESG investing, but not the WTP and return-belief questions we study.

[2] Prior work documents individual aspects of this pattern. Pastor et al. (2025a) show that institutions hold greener portfolios. Dahlquist and Ibert (2024) and Edmans et al. (2026) show that mandates and other constraints weaken the link between managers' beliefs and portfolios. Observing stated tastes, beliefs, allocations, and outside pressure jointly, we provide a more complete picture of the institutional green tilt.

relate to different real-world data. We also distinguish beliefs about greenness itself from beliefs that can also reflect other characteristics of real-world portfolios.[3]

Our factorial design also lets us measure how tastes vary across investors one characteristic at a time, including for investors who are hard to find in human samples, such as women under 35 who are strong Republicans. Several of our cross-sectional patterns match prior evidence from human surveys: political views strongly predict sustainability preferences, women tend to express stronger such preferences, institutional demand varies across types and regions, and household demand weakens in downturns (Hong and Kostovetsky, 2012; Bauer et al., 2021; Barber et al., 2021; Döttling and Kim, 2024; Giglio et al., 2025; Aron-Dine et al., 2025). This agreement lends credibility to the synthetic responses.

Our third contribution is methodological: we develop a mixing-and-reconstruction design for studying responses to historical information environments. Human surveys conducted today cannot credibly recover how investors would have responded in the past. Recent work uses fixed LLMs to elicit historical expectations under changing information environments (Bybee, 2025; Hansen et al., 2026; Wu et al., 2025; Graham et al., 2026), but a fixed model may still draw on later events in its training data (Sarkar and Vafa, 2024; Levy, 2026). Masking dates alone may leave the historical period recognizable (Lopez-Lira et al., 2026). Chronologically restricted models avoid future training information (He et al., 2025; Kelly et al., 2026), but using different models across historical dates makes it difficult to separate changes in the information environment from changes in the model itself. We instead hold the model fixed, mask identifying temporal cues, mix information modules across years, and reconstruct responses for each historical information environment from the estimated module effects. This design allows us to isolate variation in the information environment and trace how model-simulated tastes and beliefs evolve across historical environments.

Section 2 discusses the advantages and limitations of synthetic surveys. Section 3 describes the survey design. Sections 4 through 6 present the cross-sectional results. Section 7 compares synthetic responses with a published human survey and examines their reliability across runs and models. Section 8 presents the time-series analysis. Section 9 concludes.

[3]The matched-versus-unrestricted belief distinction builds on Edmans et al. (2026), who distinguish beliefs about environmental and social performance from beliefs about correlated firm characteristics. We extend this distinction to the portfolio level and elicit both belief measures separately from tastes.

2 Synthetic versus Human Surveys

This section evaluates synthetic surveys as a complement to human surveys and places our approach in the emerging literature that uses LLMs as survey respondents. Section 2.1 discusses the advantages and limitations of the approach, and Section 2.2 relates our design to the studies closest to it. Synthetic surveys offer greater experimental control, access to counterfactual environments, broader coverage, and low-cost replication, but raise concerns about external validity and model dependence. We examine these concerns in Section 7, which begins with a comparison against a published human survey.

2.1 Advantages and Limitations

Human surveys provide direct evidence on preferences and beliefs that market data do not reveal, but they face several well-known challenges, as reviewed by Stantcheva (2023). Samples are shaped by coverage, nonresponse, self-selection, and attrition, while responses can be affected by attention, question wording and order, social desirability, and perceived researcher expectations. Careful survey design can mitigate these problems but not eliminate them. Synthetic surveys relax some of these constraints while introducing others.

A synthetic survey offers five advantages for our setting. First, respondent characteristics can be varied independently. In human samples, characteristics such as income and financial literacy are correlated, and rare combinations, such as high income with low literacy, are difficult to observe in meaningful numbers. A factorial synthetic design breaks these correlations by construction, so one characteristic can change while the others remain fixed.

Second, synthetic respondents can be surveyed in counterfactual environments with their identities held fixed. These counterfactuals can vary current economic conditions, such as a recession or a boom, or the information available in a historical period. Historical counterfactuals are especially difficult for human surveys. Because each model call starts afresh, we can add, remove, or vary specific content in a synthetic respondent's information environment. Human respondents cannot be reset in this way. Someone surveyed today cannot forget what happened after 2014 or recover the attitudes they held at the time.

Third, synthetic surveys are not constrained by respondent availability or participation. Financial professionals are difficult to recruit; a recent survey of CFA-certified professionals obtained a response rate of only 3.6% (Bauer et al., 2026). Synthetic identities can represent

such hard-to-reach groups using the patterns the model has learned from human-generated data, without requiring members of those groups to participate directly. This also removes participation selection at the elicitation stage. The composition of the synthetic sample is fixed by design rather than by who has the time, interest, or willingness to respond, so differential response rates and attrition do not alter the sample.

Fourth, question modules can be isolated from one another. In the main survey, we administer each module in a fresh model call, without questions or answers from other modules, so one response cannot directly influence another.[4] Human surveys can randomize question order, but respondents may still carry information from earlier questions into later answers (Stantcheva, 2023). Separate calls also make repeated elicitation possible, allowing us to measure response noise.

Fifth, synthetic surveys make large designs feasible at relatively low marginal cost (Horton et al., 2026). Human surveys must limit questionnaire length and the number of conditions assigned to each respondent because longer surveys increase fatigue and attrition. Synthetic identities do not face these constraints, allowing every identity to complete all assigned questions and conditions rather than dividing the design across respondents. In our setting, this makes it feasible to run 104,400 model calls while preserving a common design across identities, models, and conditions. Synthetic surveys are also more replicable because other researchers can rerun the same questions on the same model specification, whereas the same human respondents cannot be observed again under identical conditions.[5]

These advantages come with three main limitations. First, synthetic responses are not direct observations of human behavior. A model conditioned on an investor description generates responses from patterns learned in its training data, and whether those responses match what people with the same characteristics would report is an empirical question (Argyle et al., 2023; Horton et al., 2026). An elicited response is a model's completion of a prompt, not a human belief (Nagel et al., 2026). Persona-prompted responses can be less dispersed than human responses (Bisbee et al., 2024). Human-feedback tuning can concentrate a group's responses around its modal position (Santurkar et al., 2023), and models may characterize some groups more from how others describe them than from how members of those groups describe themselves (Wang et al., 2025). Responses may also lean toward the view implied

[4]Bhagwat et al. (2026) similarly evaluate each persona–news combination in a separate model call.

[5]Replication is not exact when the underlying model itself changes, a limitation we return to below.

by the question (Anthis et al., 2025). On ESG questions in particular, Li and Yu (2025) find that LLM responses are more favorable to ESG than those of human investors.

Our design mitigates, but does not eliminate, these concerns in three ways. Our standing instructions ask for the respondent's own view, including any disagreement with expert or consensus opinion. Our main analyses emphasize comparisons within the same model, which remove additive response tendencies common to all identities, although biases that vary across identities can remain. We also benchmark the synthetic responses against a published human survey in Section 7.1. They reproduce the signs and cross-sectional orderings in the human data more closely than the response levels, so we place greater weight on those comparative patterns when interpreting our results.

Second, results can depend on the model and how the questions are posed. Different model versions may produce different response levels, and even the same proprietary model can give different answers to an unchanged prompt (Bisbee et al., 2024). Responses can also depend on prompt details, including wording, tone, numerical anchors, and the order of answer labels (Bisbee et al., 2024; Nagel et al., 2026). We assess these concerns in Section 7 by examining repeated runs and comparing results across model versions. Section 7.3 also examines question-order effects in our design. We emphasize patterns that recur across runs, models, and question orders over model-specific response levels.

Third, synthetic surveys relocate rather than eliminate selection. Human respondents select into surveys through availability, interest, and willingness to participate, whereas models inherit selection from the text on which they are trained. Public text may disproportionately reflect people who choose to write (Horton et al., 2026), so synthetic responses may not always represent the population distribution of the corresponding investor type. The human benchmark in Section 7.1 provides a check on this concern. Our main analysis also emphasizes differences across identities within a common design, which partly differences out components of training-data selection that are common across identities.

The emerging synthetic-survey literature provides encouraging but mixed evidence on this approach (Argyle et al., 2023; Bisbee et al., 2024; Santurkar et al., 2023; Ashokkumar et al., 2026). More importantly, alignment with human responses varies across settings (Santurkar et al., 2023; Ashokkumar et al., 2026), making it important to validate the approach in the setting we study. We do so in Section 7.1.

2.2 Related Literature

Among studies that use LLMs as survey respondents, Li and Yu (2025) are closest to us in topic. They adapt questions from the fund-manager survey of Edmans et al. (2026) and the retail-investor survey of Giglio et al. (2025), and compare the models' own ESG views with the corresponding human benchmarks. Their retail and professional exercises, however, draw on separate surveys, and their design does not systematically vary investor characteristics within either group. We instead place retail and institutional identities within a common survey design and separately elicit green tastes, return beliefs, allocations, and motives at the portfolio level from the same identities. We also vary investor characteristics, sustainability dimensions, economic conditions, and historical information environments. This design allows us to study both cross-sectional and historical heterogeneity, identify settings in which tastes and beliefs move differently, and relate both directly to portfolio choice.

Methodologically, Bhagwat et al. (2026) are closest to us in their controlled variation of investor identities. They construct 216 investor personas by crossing demographic characteristics with political orientation, ask each persona to respond to firm news with a buy, hold, or sell decision, and use dispersion across personas to measure investor disagreement. They validate this measure against human benchmarks and relate it to trading volume. Like them, we vary investor characteristics and administer persona-specific questions in separate model calls. Our design differs in both the objects elicited and the scope of the variation. Rather than using trading responses to news to measure disagreement, we separately elicit tastes, beliefs, allocations, and motives from both retail and institutional identities, so we can study how tastes and beliefs differ across investors and how both relate to portfolio choice. We also extend the controlled variation beyond investor characteristics to sustainability dimensions, economic conditions, and historical information environments.

A related identity-based design is Fedyk et al. (2026), who administer the same investment questionnaire to human participants and GPT-4 agents seeded with demographic characteristics. Their focus is the human–AI comparison, while ours is heterogeneity across synthetic investors and environments.

Another methodological comparison is with studies that use LLMs to simulate survey responses at past dates. Hansen et al. (2026) recreate the Survey of Professional Forecasters from 1999 to 2023 by combining participant personas with information available at each

survey date. Graham et al. (2026) use CFO profiles and prior survey answers to predict individual CFO responses, while Wu et al. (2025) use LLMs to extend a household survey back to 1990. All three studies elicit expectations rather than tastes, and none relates the responses to portfolio choice. Their prompts also state the survey date. Even with instructions to ignore later information and checks for leakage, a model that is told the date may still draw on what it has learned about later years. We instead mask identifying temporal cues, mix information modules across years, and estimate their effects before reconstructing responses for each historical information environment. Our historical responses are therefore recovered from variation across mixed information environments rather than elicited directly from prompts corresponding to a particular year.

Nagel et al. (2026) raise a more direct challenge to treating LLM responses as human beliefs. In prediction markets and analyst forecasting, they show that LLM probabilities can differ from human beliefs in both levels and implied economic relationships. These discrepancies are sensitive to prompt details and training exposure, and in prediction markets they persist after averaging over demographic personas. Matching average response levels to human benchmarks does not recover the relationships observed in human data.

Their setting differs from ours. They study LLM probabilities for binary events whose outcomes are later observed, whereas we separate tastes from beliefs and study how each relates to portfolio choice. This difference does not exempt our design from their critique: every response we elicit is a model completion of a prompt. Our design responds to their concerns in three ways. First, we do not interpret synthetic response levels as population estimates of human preferences or beliefs. Section 7.1 instead asks whether our synthetic retail belief responses reproduce the signs and cross-sectional orderings observed in a published human survey, and we place greater weight on the patterns that do. Second, we assess model and prompt dependence directly: Section 7 compares repeated runs and model versions, while Section 7.3 examines question-order effects. We administer each module in a separate model call, so the answer to one module never appears in the prompt for another, and the relation between tastes and beliefs cannot reflect one answer anchoring the other. Third, the training-exposure concern motivates our historical design. Masking and cross-year mixing reduce the correspondence between any individual call and a recognizable historical period, although they cannot remove information already learned by the model. These checks do

not make synthetic responses equivalent to human responses, but they discipline which comparative patterns we emphasize in interpreting the results.

3 Survey Design

Our survey studies how green tastes, return beliefs, portfolio allocations, and investment motives vary with investor characteristics, the sustainability dimension, and economic conditions. The 480 retail identities form a complete grid over political affiliation, income, age, gender, and financial literacy. Age and income categories group the bins of the Survey of Consumer Finances (SCF) (Board of Governors, 2023). We describe political affiliation through party registration and donation behavior. The 60 institutional identities form a complete grid over institution type, size, and geography, and each answers as the institution's investment decision-maker. We present each identity to the model as a short paragraph stating these characteristics, an approach known as persona prompting. Details are in the Appendix.

Each identity completes five modules under six conditions: an aggregate-ESG score in a normal economy, which we call the baseline; separate environmental, social, and governance scores; and aggregate ESG in a recession or a boom. We impose no knowledge date, so each model uses the information available in its training. A cell is one identity under one condition. We administer the five modules in separate model calls, each containing two questions and no questions or answers from other modules.

The five modules elicit four objects from the same identity. The preference module elicits WTP for greenness in percentage points of annual return. Two belief modules elicit expected returns, in the same units, on a matched pair of portfolios and on typical real-world portfolios. The allocation module elicits the percentage of wealth or assets held in U.S. stocks and the percentage of those stocks held in the greener matched portfolio. The motive module asks how much of the greener-portfolio holding reflects the identity's own choice rather than outside pressure. We describe the questions where their answers are first analyzed: tastes in Section 4, beliefs in Section 5, and allocations and motives in Section 6.

The main model, ChatGPT-5.6 Sol, answered every cell three times. Three additional models—ChatGPT-5.5, Claude Opus 5.0, and Claude Fable 5—answered every cell once. We call these four the full-grid models. Nine further model versions answered the preference and belief questions on fixed subsamples. The design comprises 104,400 model calls (Table 1).

Appendix A reproduces all survey instruments, including identity descriptions, question modules with their value and economy variants, the validation questionnaire, and dated-sample instruments. The replication package provides model identifiers and archived calls.

4 Green Tastes

This section examines how green tastes, measured by willingness to pay for greenness, vary across investors and economic conditions.

The preference question describes two diversified U.S. equity portfolios, H and L, that differ only in their sustainability scores. Portfolio H scores at the 80th percentile and Portfolio L at the 20th percentile of a fixed greenness scoring methodology. Every other financial and non-financial characteristic, including industry mix, cash flows, risk, size, and fees, is held fixed now and for the next ten years.[6] Each identity reports the average annual return it would give up to hold Portfolio H and the additional return it would require to hold Portfolio L. Answers may be positive, zero, or negative, allowing investors who dislike greenness to demand compensation for holding Portfolio H.

Let g denote the annual return given up for Portfolio H and r the extra annual return required for Portfolio L, both in percentage points of annual return over ten years. We define *willingness to pay* (WTP) as their average:

$$\text{WTP} \;=\; \frac{g+r}{2}. \tag{1}$$

4.1 Retail Investors

Political affiliation is the strongest determinant of green tastes across the 480 retail identities. In Figure 1, average WTP in the main model falls steadily from 1.04% per year for strong Democrats to 0.02% for strong Republicans, or about a quarter of a percentage point per step on the five-point political scale.

[6] Brodback et al. (2026) vary charitable benefits while holding assets' financial risk-return profiles fixed. Their experiment elicits an incentivized asset-price premium; we elicit a hypothetical annual-return tradeoff.

This ordering is also apparent in the distribution across identities. WTP is positive for 92% of the 480 identities. The exceptions are almost all strong Republicans: of the 38 identities with a negative WTP, 37 are strong Republicans.

All four full-grid models produce the same steep and steady decline in WTP from strong Democrats to strong Republicans (Figure 1). The 95% intervals of neighboring political levels do not overlap in any model. In three of the four models, the average WTP of strong Republicans is negative, reaching -0.37 in Claude Opus 5.0.

To rank the five retail attributes, we regress the cell-level WTP on category indicators for all of them at once (Table 2). Since the grid is a complete factorial, the attributes are uncorrelated by construction, so ΔR^2, the drop in R^2 from deleting an attribute, measures the share of the variation that attribute accounts for. Politics has a ΔR^2 of 0.75, compared with no more than 0.03 for any other attribute and less than 0.08 for income, gender, financial literacy, and age combined. It ranks first in all 13 models.[7] The other characteristics have smaller but statistically significant effects on WTP: women state a WTP 0.12 percentage points per year higher than men, the top income group 0.19 points higher than the bottom, low-literacy identities 0.11 points higher than high-literacy ones, and the youngest group 0.09 points higher than the oldest.[8]

Income and financial literacy have opposing conditional effects. Holding literacy fixed, WTP rises with income; holding income fixed, WTP falls with literacy. In observational data, higher-income investors tend to be more financially literate (Lusardi and Mitchell, 2014), making these effects difficult to distinguish. The factorial grid separates them by varying income and literacy independently.

[7]The direction of the political difference matches evidence on human investors. Mutual fund managers who donate to Democrats hold fewer socially irresponsible stocks (Hong and Kostovetsky, 2012), and pension participants who vote for parties with a sustainability-focused agenda are more likely to support sustainable investment by their fund (Bauer et al., 2021). Investors in Democratic-leaning areas more often cite ethical motives for holding ESG funds (Giglio et al., 2025). A German household survey also finds that signed annual-interest tradeoffs for green deposits vary mainly with political views rather than wealth or other demographics (Aron-Dine et al., 2025). Those hypothetical same-bank deposit choices measure a convenience yield on safe assets, whereas our WTP is an annual-return tradeoff on matched equity portfolios.

[8]The gender difference is consistent with survey evidence that women state stronger sustainability preferences (Bauer et al., 2021; Giglio et al., 2025).

4.2 Institutional Investors

Institutional identities state a positive but substantially lower WTP than retail identities: 0.18% per year at baseline in the main model, compared with 0.49% for retail.[9] Based on ΔR^2 again, institution type explains 0.62 of the variation across the 60 institutional identities, geography explains 0.25, and size 0.02 (see Figure 2 and Table 3, the institutional counterpart of Table 2). Endowments and foundations state a WTP of 0.30%, pension funds 0.20%, and hedge funds and mutual fund companies 0.12%.

European institutions state the highest WTP, at 0.24%, and U.S. institutions the lowest, at 0.12%. This geographic ordering is consistent with the conditional comparison of Pastor et al. (2025a), in which European 13F institutions hold greener U.S. equity portfolios than U.S. 13F institutions.[10]

4.3 The Two Sides

The two components of WTP in equation (1) are asymmetric. Across all 3,240 retail and institutional cells and all score and economy conditions, identities in the main model would give up 0.33% per year to hold Portfolio H but require 0.52% to hold Portfolio L. Required compensation exceeds the return sacrifice in every model (Figure B.1). The same pattern holds within each investor class in every model: retail identities give up 0.35% and require 0.56% in the main model, while institutions give up 0.14% and require 0.20%.

One possible interpretation is loss aversion around a moral reference point: if a typical portfolio's sustainability score is the reference, falling below it registers as a loss, and losses weigh more than gains of the same size.[11]

[9]The low institutional WTP is consistent with survey evidence on fund managers: only 27% of those surveyed by Edmans et al. (2026) would tolerate a company sacrificing even one basis point of long-term risk-adjusted return to improve its environmental and social performance. In contrast, Baker et al. (2024) estimate that in 2020–2021 institutional investors were willing to pay a higher fee than retail investors for index funds with an ESG mandate. In their 2023 estimates, the premium is negative for both groups, slightly more so for institutions.

[10]In their study of limited partners in venture capital funds, Barber et al. (2021) likewise estimate a high willingness to pay for impact among foundations and European investors.

[11]A human experiment finds a similar asymmetry, with negative externalities affecting investment choices about three times as much as equally sized positive externalities (Humphrey et al., 2023). In contrast, Bonnefon et al. (2025) find approximately symmetric valuations of positive and negative externalities.

4.4 States of the Economy

A recession lowers the retail WTP, significantly in three of the four models, but leaves the institutional WTP nearly unchanged. The retail WTP is procyclical: relative to a normal-times level of 0.49%, the main model's retail identities lower it by 0.08% in a recession and raise it by 0.02% in a boom. All four models state their lowest retail WTP in a recession (Figure 3, Panels A and C). The pattern is consistent with households treating non-pecuniary values as a normal good (Bansal et al., 2022; Döttling and Kim, 2024): the willingness to sacrifice returns declines, but does not disappear, when economic conditions deteriorate.[12]

Institutional WTP is much less sensitive to economic conditions (Figure 3, Panels B and C). The largest institutional deviation from normal times in any model or state is 0.02%; the main model's recession effect is +0.01%, compared with −0.08% for retail. The motive responses in Section 6.3 suggest a possible explanation: institutions attribute about half of their Portfolio H holdings to outside pressure, and their mandates and client obligations may be more persistent across the business cycle than households' tastes.

5 Beliefs

The next two modules elicit expected returns. The matched-belief module retains the matching language of the preference question in Section 4 and asks for the expected annual returns over the next ten years on Portfolios H and L. The unrestricted-belief module drops the matching restriction and asks the same question about typical real-world portfolios at the 80th and 20th percentiles of the sustainability score, which we denote by $\tilde{H}$ and $\tilde{L}$. We define the *matched belief gap* and the *unrestricted belief gap* as the corresponding differences in expected returns:

$$\begin{aligned} \text{Matched belief gap} &= \mathrm{E}(R_H) - \mathrm{E}(R_L), \\ \text{Unrestricted belief gap} &= \mathrm{E}(R_{\tilde{H}}) - \mathrm{E}(R_{\tilde{L}}), \end{aligned} \tag{2}$$

where $\mathrm{E}(R_P)$ denotes the identity's expected annual return on portfolio P.

[12] Investors kept their focus on sustainability in the 2020 crash (Pastor and Vorsatz, 2020), and retail flows into high-ESG funds declined while institutional flows did not (Döttling and Kim, 2024). Flows during a short shock are a different object from a stated long-term sacrifice of returns under a recession.

A positive matched belief gap means that the identity expects Portfolio H to outperform Portfolio L. Since the two portfolios differ only in their sustainability scores, an identity with a nonzero matched gap attributes an expected-return effect to greenness itself. A positive unrestricted belief gap means that the identity expects typical real-world high-score portfolios to outperform typical low-score portfolios. Unlike the matched gap, the unrestricted gap can also reflect differences in sector composition, growth, valuations, and other characteristics that accompany the relevant sustainability dimension in practice.

5.1 Retail Investors

Both belief gaps are strongly ordered by political affiliation (Figure 4, Panel A). Strong Democrats expect Portfolio H to outperform Portfolio L by 0.47% per year, whereas strong Republicans expect Portfolio H to underperform by 0.52%. One possible interpretation is that identities project their preferences onto their beliefs, or believe that other investors share those preferences and will move prices in the same direction.

More financially literate identities are more likely to expect Portfolios H and L to earn exactly the same return. In the main model, 47% of high-literacy identities do so, compared with 4% of medium-literacy identities and none of the low-literacy identities.

The unrestricted gap follows the same ordering but is more pronounced, ranging from $+0.95\%$ for strong Democrats to -1.08% for strong Republicans. This amplification means that beliefs about the characteristics accompanying greenness in real-world portfolios reinforce, rather than offset, the political differences already present in beliefs about greenness itself.

The two belief gaps have similar means of essentially zero but different dispersions. The cross-identity standard deviations are 0.49% and 0.90% for the matched and unrestricted gaps, respectively (Figure 4, Panels B and C). In both distributions, political affiliation produces a clear ordering: moving from strong Democrat to strong Republican shifts the distribution steadily toward more negative beliefs. Dropping the matching restriction therefore widens political disagreement substantially without materially changing the average belief gap.

5.2 Institutional Investors

In the main model, all 60 institutional identities expect Portfolios H and L to earn exactly the same return, meaning their matched belief gap is zero. When greenness is the only stated difference between the two portfolios, institutions assign no expected-return effect to it.[13]

The pattern changes once the matching restriction is removed. Institutions expect typical real-world high-sustainability-score portfolios to underperform low-score ones by 0.35% per year on average, and the unrestricted belief gap is negative for 56 of the 60 identities. They therefore expect the characteristics that accompany greenness in real-world portfolios to lower expected returns, even though they state a positive WTP for greenness itself.

The unrestricted gap also varies systematically by institution type and geography (Figure 5). Hedge funds are the most pessimistic about real-world high-score portfolios, with an average gap of -0.53%, while insurance companies are the least pessimistic, at -0.25%. U.S. institutions are also more pessimistic than European institutions (-0.43% versus -0.23%). The geographic ordering is the same for tastes and beliefs: European institutions state both stronger green tastes and less negative return beliefs than U.S. institutions.

For typical real-world portfolios, the institutional unrestricted belief gap is 0.31 to 0.42 percentage points lower than the retail gap across the four full-grid models.[14] For the matched portfolios, we find no evidence that institutions expect higher returns from greenness itself than retail identities do: the institutional matched belief gap is lower than the retail gap in three of the four models and not significantly different from it in the fourth model. This matched comparison is the one that matters for the allocation question in Section 6, because that question uses the same two portfolios.[15]

[13]The exact zero is a feature of the main model. ChatGPT-5.5 reports a zero institutional matched gap in 82% of the 60 cells, while the average gap is -0.38 in Claude Opus 5.0 and $+0.07$ in Claude Fable 5.

[14]Human survey evidence does not allow a direct comparison. Fund managers surveyed by Edmans et al. (2026) are mostly optimistic about firms with good environmental and social performance, while retail investors surveyed by Giglio et al. (2025) on average expect ESG portfolios to underperform the market, but the two surveys ask different questions of different samples.

[15]Since the main survey imposes no knowledge date, these comparisons describe beliefs in the current information environment. In the reconstructed historical environments, retail and institutional beliefs about typical real-world green portfolios are close, while institutional WTP remains lower (Figure B.8).

5.3 Dimensions of Value

Sustainable-investing theory often summarizes greenness with a single score, leaving open which underlying dimensions investors value most (Pastor et al., 2025b). We compare environmental, social, and governance scores within the same 540 identities, including 480 retail and 60 institutional identities. Each identity answers the same questions with the sustainability score defined in turn by the environmental, social, or governance dimension rather than aggregate ESG. Figure 6 reports the resulting changes in WTP (Panel A) and in the matched belief gap (Panel B), pooled over the 540 identities.

We find that tastes change little when the score measures only environmental or social characteristics. In every model, the change in WTP from the aggregate-ESG baseline is no larger than 0.15 percentage points for either dimension. Governance produces less consistent responses: the two ChatGPT models lower WTP by 0.08 and 0.06 points, while the two Claude models raise it by 0.49 and 0.16 points (Figure 6, Panel A).

Beliefs distinguish the three dimensions more consistently. Governance stands out: in every model, respondents on average expect high-governance portfolios to outperform otherwise identical low-governance portfolios. Replacing aggregate ESG with a governance-only score raises the matched belief gap by 0.39 to 1.09 percentage points (Figure 6, Panel B). The increase appears among both retail and institutional identities.[16] The environmental dimension has little effect on beliefs, while the social dimension has a more modest positive effect, reaching at most 0.29 points.

Respondents thus attach greater financial importance to governance than to environmental or social characteristics.[17] Since Portfolios H and L differ only in their governance scores, the higher expected return on Portfolio H suggests that respondents treat governance quality itself as informative about firm value and future performance.

[16]The pooled results are weighted toward retail identities by construction (480 of the 540). Figure B.2 reports the two investor classes separately. In both classes, governance produces the largest increase in the matched belief gap in all four models.

[17]Fund managers surveyed by Edmans et al. (2026) likewise rank governance above environmental and social performance as a driver of long-term firm value.

6 Portfolios

The final two modules elicit hypothetical portfolio allocations and investment motives.

The allocation question uses the same matched Portfolios H and L and the same ten-year horizon as the preference and matched-belief questions. It asks how the identity would allocate its capital. First, we ask each identity what share of its assets—financial wealth for retail and assets under management (AUM) for institutions—it would invest in the U.S. stock market. Second, we ask what share of the stock allocation it would place in Portfolio H, with the remainder in Portfolio L. The first answer defines the *participation margin*, while the second defines the *composition margin*. Since Portfolios H and L are the only two equity choices, a Portfolio H share of 50% represents an even split. This share measures the identity's tilt toward the higher sustainability score rather than an actual green holding.

The motivation question assumes that the identity holds some of Portfolio H and asks how much of that holding reflects its own choice versus outside pressure. For retail identities, outside pressure can come from friends, family, colleagues, or others whose opinions matter to them. For institutional identities, it can come from mandates, the investor base, trustees or clients, regulatory or industry expectations, or reputational considerations. The own-choice and outside-pressure shares sum to 100%.

6.1 Participation and Composition

Compared to retail identities, institutions place a smaller share of their assets in U.S. equities but tilt more toward Portfolio H (Table 4). Retail identities invest 52% of their financial wealth in stocks and allocate 60% of that stock investment to Portfolio H. Institutions invest 23% of AUM in stocks but allocate 70% of those stocks to Portfolio H. Multiplying the two shares within each response and then averaging gives an overall Portfolio H holding of 32% of financial wealth for retail identities and 16% of AUM for institutions.

For retail identities, the two margins vary with different characteristics. Income explains 48% of the variation in the stock share and age explains 32%; politics explains less than 1% (Table B.1). The bottom income group invests 17 percentage points less of its wealth in stocks than the middle-income group, the oldest group 11 points less than the middle-aged group, and women 3 points less than men. Politics, by contrast, dominates the composition margin, with $\Delta R^2 = 0.93$ for the Portfolio H share.

For institutional identities, institution type matters for both margins. It explains most of the variation in the U.S. equity share ($\Delta R^2 = 0.74$, compared with 0.11 for geography). Type and geography both matter for the Portfolio H share, with ΔR^2 of 0.55 and 0.38, respectively. AUM size explains little of either margin (Table B.2).

6.2 Green Tilts, Tastes, and Beliefs

Political affiliation has a much larger effect on portfolio composition than on stock-market participation. Strong Democrats allocate 85% of their stock investment to Portfolio H, compared with 29% for strong Republicans, with a monotonic decline across the political scale in all four models (Figure 7, Panel A). This pattern mirrors the political gradients in tastes and beliefs documented in Sections 4.1 and 5.1. In contrast, the share of wealth invested in stocks varies little across the political scale (Panel B).

The Portfolio H share is strongly associated with WTP across the 480 baseline retail identities (Figure B.3). The two answers come from separate model calls, yet their correlation is 0.83. Every Democratic and Neutral identity allocates more than half of its stocks to Portfolio H, whereas no strong-Republican identity does. Across all 3,240 cells, the same relation appears in every model (Table 5): one percentage point of WTP is associated with a 21-to-45-percentage-point increase in the Portfolio H share. WTP alone accounts for 48% to 79% of the variation in portfolio composition across cells, depending on the model.

Green tilts relate to both tastes and beliefs, but more strongly to tastes. When WTP and the matched belief gap in equation (2) enter the same regression, both are positively and significantly associated with the Portfolio H share (Table 5).[18] As both are measured in annual percentage points, their coefficients are directly comparable. The WTP coefficient is larger in all four models; in the main model, it is 26.3, about twice the matched-belief coefficient of 13.7. Adding the matched belief gap to WTP raises R^2 from 0.50 to 0.57.

[18]Evidence on human investors also links green holdings to both non-pecuniary motives and return beliefs. Riedl and Smeets (2017) find that social preferences, more than financial motives, explain holdings of socially responsible funds. Giglio et al. (2025) find that ESG holdings rise with both expected ESG returns and ethical motives.

6.3 Own Choice versus Outside Pressure

Institutions state weaker green tastes than retail identities (Figure 8, Panel A), yet they tilt more strongly toward Portfolio H. They allocate about 70% of their equity investment to Portfolio H in the main model, compared with 60% for retail identities (Table 4), and the institutional tilt is larger in each of the four models (Figure 8, Panel C). This ordering is also consistent with actual holdings: the AUM-weighted aggregate portfolio of 13F institutions tilts green, while the combined non-13F residual, which includes households and smaller institutions, tilts brown (Pastor et al., 2025a). Return beliefs about Portfolios H and L do not explain the stronger institutional tilt, because we find no evidence that institutions expect Portfolio H to outperform L by more than retail identities do (Section 5.2).

Instead, the motivation responses help explain this contrast. Retail identities attribute 93% of their Portfolio H holding to their own choice in the main model and at least 81% in every model. In contrast, institutions attribute roughly half or more of their Portfolio H holding to outside pressure in every model (Figure 8, Panel D; Table B.4).

Within institutions, the own-choice share broadly follows WTP across institution types: types with higher stated WTP also tend to attribute more of their Portfolio H holding to their own choice (compare Figures 2 and B.4). This relation is descriptive, since own choice can reflect expected-return beliefs as well as non-pecuniary preferences.

The stronger institutional tilt remains after we control for tastes and for beliefs about Portfolios H and L. Adding an institution indicator to the regressions of Table 5 that include both WTP and the matched belief gap yields a coefficient of 19.2 percentage points in the main model and 12 to 24 points across the four models (Table B.3, Panel C).

A stronger green tilt therefore does not necessarily indicate stronger green tastes. It can also reflect return beliefs or, as here, outside pressure.

7 Reliability and Validation

The preceding sections document systematic differences in tastes, beliefs, allocations, and motives across identities and conditions. In this section, we first compare synthetic responses with those in a published human survey. We then examine how sensitive our results are to repeated elicitation and to the choice of model, using three comparisons: repeated runs of the

main model, results across the four full-grid models, and responses from 13 model versions on a common set of cells.

7.1 Comparison with a Human Survey

We ask whether synthetic respondents reproduce the central patterns observed in a published human survey when asked the same questions. We conduct two exercises. The first administers the published survey questions to synthetic profiles constructed to match the characteristics of the original survey sample. The second administers the same questions to profiles that preserve the joint distribution of household characteristics in the SCF. The first exercise provides a direct human benchmark. We find that the synthetic respondents reproduce the signs and cross-sectional orderings in the published data, including the expectation that ESG portfolios underperform the market, but not the dispersion of human responses. The second exercise shows that the same expectation persists when profiles reflect combinations of characteristics observed in the population.

We take the questionnaire verbatim from two published surveys, so it is independent of the design in Section 3. The three ESG questions come from Figure 1 of Giglio et al. (2025), and the market-return question from the survey appendix of Giglio et al. (2021). We add only the instructions needed to administer these questions to a model, including the respondent identity, the reference date, and the required output format (see Appendix A.3).[19]

The first exercise compares the published human sample with a matched synthetic sample. We draw 1,000 synthetic profiles to match the published sample's marginal distributions of age, wealth, gender, county political majority, and flood risk. Each profile answers all four questions in ChatGPT-5.6 Sol and Claude Fable 5. The results are in Table 6.

Two of the four questions ask for the expected annual return over the next ten years on a diversified U.S. ESG equity portfolio and on the U.S. stock market. We call the difference between the two the *ESG-minus-market belief gap*. Both models reproduce the sign of the average gap and the main cross-sectional patterns in the published data. The gap is negative in all three samples: -2.06% in the human data, -0.56% in ChatGPT, and -0.91% in Claude. The cross-sectional ordering is also similar. Women are less pessimistic about ESG returns than men, profiles from Democratic-majority counties are less pessimistic than those from

[19]We use December 2023, the final wave of the published survey, as the reference date.

Republican-majority counties, and the belief gap becomes more negative as climate concern declines. The same pattern appears across stated motives for ESG investing: respondents who report "no specific reason" as their motive for ESG investing expect the largest ESG underperformance in the human sample as well as in both models.

The synthetic responses are substantially less dispersed than the human responses. This compression is consistent with Bisbee et al. (2024), who document a similar pattern for persona-prompted survey responses. The cross-profile standard deviation of the belief gap is 0.31 in ChatGPT and 0.51 in Claude, compared with 5.34 in the human data.

One possible reason for the compression is that a model conditioned on a category tends to give the response typical of that category (Wang et al., 2025), whereas people within a category differ substantially. In the survey we benchmark against, most of the dispersion in beliefs lies within demographic groups rather than between them (Giglio et al., 2025). Another possibility is participation selection: people with stronger views may be more likely to participate when interest in the survey topic affects response rates (Groves et al., 2004; Stantcheva, 2023). Such selection would make observed human responses more dispersed than those in the underlying population. Our data cannot distinguish between these explanations.

The second exercise preserves the joint distribution of household characteristics rather than matching each marginal distribution separately. The two models answer the same four questions for profiles representing all 4,595 households in the 2022 SCF (Board of Governors, 2023). Each profile describes one household's combination of characteristics, and we aggregate the synthetic answers using SCF population weights (Table 7). We instruct the models to answer as of 2022, the SCF survey year.

Both models again expect ESG portfolios to underperform. The weighted belief gap is -0.60% (standard error 0.004) in ChatGPT-5.6 Sol and -1.11% (standard error 0.010) in Claude Fable 5. Weighted and unweighted means are nearly identical, so the result does not depend on the SCF weights. The two models also agree profile by profile: the sign of the gap is the same in 98% of household profiles.

The benchmark is limited to retail return beliefs in two models and therefore does not directly validate our taste measure, allocations, motives, institutional comparisons, or historical responses. Within its scope, however, the synthetic responses reproduce the main signs and cross-sectional orderings in the human data more closely than their levels or

dispersion. We therefore interpret our results primarily through comparative patterns, while treating absolute response levels and outcomes outside the benchmark more cautiously.

7.2 Repeated Runs

The main model answered all 3,240 identity-condition cells three times, allowing us to measure run-to-run variation directly. For WTP, the average within-cell standard deviation across the three repeats is 0.09%, compared with 0.34% across cell means. Beliefs show similar repeatability: the corresponding within-cell and across-cell standard deviations are 0.13% and 0.47% for the matched belief gap, and 0.22% and 0.86% for the unrestricted belief gap. For all three measures, the average within-cell standard deviation is about one quarter of the standard deviation across cell averages. We thus see that differences across identities and conditions are much larger than run-to-run variation.

For allocations and motives, run-to-run variation is even smaller relative to cross-cell variation. The average within-cell standard deviation is less than one fifth of the standard deviation across cell averages. The within-cell and across-cell standard deviations are 3.4% and 18.8% for the Portfolio H share, 2.5% and 16.4% for the stock share, and 2.5% and 16.5% for the own-choice share of Portfolio H holdings, respectively.

7.3 Across Full-Grid Models

The main patterns in green tastes recur across the four full-grid models. WTP declines steeply and monotonically from strong Democrats to strong Republicans in every model (Figure 1). The asymmetry between the two sides of WTP is also common across models: identities consistently require more return to hold Portfolio L than they are willing to give up to hold Portfolio H (Figure B.1). Retail WTP is also lowest in a recession in all four models, and the decline is significant in three of them (Figure 3).

The belief results are similarly consistent across models. In every model, governance raises the matched belief gap by more than the environmental or social dimension does (Figure 6, Panel B; Figure B.2).

The portfolio results are similarly robust. In every model, political affiliation strongly orders the share of stocks allocated to Portfolio H, while the share of wealth invested in stocks varies much less across the political scale (Figure 7). In every model, both WTP and

the matched belief gap are positively associated with the Portfolio H share. Both coefficients remain positive when the two variables enter the same regression, and the WTP coefficient is the larger of the two in every model (Table 5).

The institutional-retail contrast also holds across all four full-grid models (Figure 8). Institutions state weaker green tastes than retail identities, yet allocate more of their equity to Portfolio H and attribute less of that holding to their own choice. Their unrestricted belief gap, which concerns typical real-world portfolios rather than Portfolios H and L, is also lower (Panel B). All four differences are significant at the 5% level in every model. The WTP and belief differences are also robust to question order: including an order indicator changes the estimated baseline institution-retail differences by 0.01% or less in every model.[20]

7.4 Across Model Versions

In addition to the four full-grid models, we asked nine other model versions to answer the preference and belief questions on fixed subsamples. We compare all 13 models on the same 240 cells. Table B.5 lists the 13 models with their documented knowledge cutoffs.

Across these common cells, response levels vary with model version (Figure B.5). Models with documented 2023 cutoffs report the highest WTP, reaching 2.2% for GPT-4o mini, whereas models with cutoffs from mid-2024 onward report WTP between 0.2% and 0.5%. The matched belief gap follows a similar pattern: the 2023-cutoff models report gaps of 0.9% to 2.5%, while the later models report gaps between -0.2% and $+0.4\%$. The association between response levels and documented cutoff dates may reflect differences in the information available to the models, but model versions also differ in capability and other features.

Despite this variation in levels, the cross-sectional political ordering is consistent across models. The strong-Democrat-minus-strong-Republican WTP gap is positive in all 13 models, ranging from 0.4% to 1.7% (Figure B.6).

8 Variation over Time

The model-version comparisons in Section 7.4 cannot separate differences in training information from differences in model capability. To study variation over time while holding

[20]Appendix A.2 describes the fixed hash-based rule used to assign question order.

the model fixed, we instead vary the information environment over 2014–2025 for seven identities in the main model, five retail and two institutional. We present each identity with cards that combine masked information from different years and use the resulting responses to reconstruct tastes and beliefs for each year by regression. To preview our results, the reconstructed series show that tastes vary little, whereas beliefs about typical real-world green portfolios become more pessimistic after 2020.

Sections 8.1 through 8.5 describe the design, present the reconstructed series, and check the mixing and the reconstruction. Section 8.6 compares the information modules with a placebo benchmark for estimation noise. Section 8.7 relates the reconstructed series to real-world data on sustainable investing and on attention to climate and ESG.

8.1 Why Masking Alone Is Not Enough

Asking a model to answer as of a past date does not limit the information in its training data. A model asked today to answer as of 2018 may still draw on information from later years, creating information leakage and look-ahead bias (Sarkar and Vafa, 2024; Levy, 2026; Lopez-Lira et al., 2026). To study how the same identities respond to different historical information environments while holding the model fixed, we need to control what the model is shown rather than simply instructing it to answer as of a past date.

A natural way to control the information presented to the model is to mask explicit dates and other identifying details. Masking alone, however, may not prevent the model from recognizing the historical period (Lopez-Lira et al., 2026). In our setting, combinations of interest rates, disasters, and policies can remain distinctive even after dates and names are removed. Removing enough information to eliminate these clues would also strip away much of the economic environment we want to study. We therefore combine masking with cross-year mixing: each card retains substantive information, but its modules are drawn from different years and no longer describe a single historical period.[21]

To implement this cross-year mixing, we divide each year's information environment into ten modules. Nine of them vary by year and cover the economy and markets, the sustainable-investment market, public views, news and investor discussion, ownership and

[21] Engelberg et al. (2026) use iterative masking and paraphrasing to reduce the recognizability of firms and dates in financial text. Our approach instead mixes masked information modules across years.

capital allocation, climate and disasters, workplace conditions, governance and enforcement, and policy and regulation. We reserved the tenth module for firm-level environmental and social characteristics but found no point-in-time data. We therefore hold its content fixed across cards and use it as a placebo in Section 8.6. We mask the nine varying modules before mixing by removing years and dates, names that identify an administration, and combinations of events that would reveal a single historical year.

We then combine the modules across years to form mixed information cards. The nine year-varying modules and 12 source years generate 12^9 possible combinations, from which we select a balanced set of 288 cards. Each module-year pair appears on exactly 24 cards, and each card draws from at least four different years, with at most three modules from any one year. For example, a card may combine economic and market information from 2016, public views from 2021, and climate and disaster information from 2019. These restrictions balance the coverage of module-year pairs and prevent any one year from dominating a card.

8.2 Elicitation under Mixed Information

We administer the same preference and belief questions as in Sections 4 and 5, preceded by a mixed information card. Relative to the main survey, we change only two of the standing instructions that precede every question. First, we replace the economy-state description with an instruction to treat all information on the card as simultaneously true and available at the time of the decision, without assuming that the stated conditions persist over the full ten-year horizon. Second, whereas the main survey lets the respondent draw on the model's general knowledge, we instruct the respondent to base its answer on the card together with its own values and judgment, without using outside knowledge to add to, correct, or reconcile the information provided. Appendix A.4 reproduces the modified instructions, the card structure, the year-guessing probe, and the opening of one example card.

8.3 Reconstructing Historical Responses

The mixed cards do not correspond to actual historical years. The purpose of the reconstruction is to use responses to these mixed environments to estimate how an identity would respond if all nine year-varying modules came from the same year.

Seven identities answer all 288 mixed cards: the five political affiliations at a fixed retail profile (age 45–64, middle income, male, high financial literacy), a large U.S. mutual fund company, and a large U.S. pension fund. For each identity and outcome, we regress the 288 card-level responses on a full set of module-year indicators. There is one indicator for each of the 108 combinations of nine varying modules and 12 source years. For any given card, exactly nine indicators equal one, one for each module and its assigned source year:

$$\text{answer}_c \;=\; a \;+\; \sum_{m=1}^{9} \sum_{t=2014}^{2025} b_{m,t} D_{m,t,c} \;+\; e_c, \tag{3}$$

where $D_{m,t,c}$ equals one if module m on card c is drawn from year t, and zero otherwise. Each coefficient $b_{m,t}$ therefore corresponds to one module-year pair. Since all nine sets of module-year indicators enter the regression together, we estimate each coefficient conditional on the year assignments of the other eight modules. Each module-year pair appears on 24 cards, alongside different combinations of years for the other modules, providing the variation used to estimate these coefficients. The design is summarized in Table B.6.

For each module, exactly one of its 12 year indicators equals one on every card. The absolute levels of the 12 coefficients for that module cannot be distinguished from the intercept: adding the same constant to all 12 coefficients and subtracting it from the intercept would leave the fitted values unchanged. We normalize the 12 coefficients within each module to an average of zero. Under this normalization, $b_{m,t}$ measures the contribution of module m in year t relative to that module's average contribution across the 12 years. The tenth module does not vary across cards and is absorbed into the intercept.[22]

Having estimated a coefficient for every module-year pair, we use these coefficients to reconstruct the response for each historical year. For year t, we take the nine coefficients associated with year t, one from each varying module, and combine them with the intercept:

$$\widehat{\text{answer}}(t) \;=\; a \;+\; \sum_{m=1}^{9} b_{m,t}. \tag{4}$$

Although no mixed card contains all nine modules from the same historical year, we estimate the coefficient for each module-year pair from the 24 cards on which that pair appears. We

[22]We estimate the coefficients using ridge regression, with the penalty chosen by cross-validation, to limit overfitting given the number of module-year effects relative to the 288 cards.

then combine the nine coefficients for year t to reconstruct the response for that year's information environment.

Finally, we aggregate the seven reconstructed identity-level series using annual corporate-equity holdings from the Federal Reserve's Z.1 accounts (Board of Governors, 2026). We map the household sector to the five retail identities, the mutual fund and ETF sector to the mutual fund identity, and the pension sector to the pension identity. We use the three sectors' annual corporate-equity holdings to construct weights that sum to one each year, with the household-sector weight divided equally across the five political identities.[23]

These three sectors account for an average of 77.1% of the corporate-equity holdings recorded in the Financial Accounts over 2014–2025. The seven identities also capture the main cross-sectional variation documented earlier. The five retail identities differ only in political affiliation, which is the strongest retail predictor of WTP and of both belief gaps. The two institutional identities represent sectors that together account for an average of 37.7% of corporate-equity holdings over the same period.[24] With the annual Z.1 holdings weights, the aggregate series reflect both the differences across these investor classes and their changing relative importance over time. We use the resulting series to trace broad movements in tastes and beliefs across historical information environments.

8.4 The Reconstructed Series

Figure 9 shows the time series of reconstructed tastes and beliefs. The aggregate unrestricted belief gap becomes more negative after 2020, indicating that respondents come to expect lower returns on typical real-world green portfolios relative to brown ones. The gap falls from close to zero in 2020 to -0.25% in 2023 and stands at -0.19% in 2025. Giglio et al. (2025) document a similar decline over a shorter window: among the Vanguard investors they survey, the average expected return on ESG portfolios relative to the market fell from -0.95% to -2.56% between June 2021 and December 2023.

Tastes are much more stable in comparison. The aggregate reconstructed WTP remains between 0.22% and 0.25% per year over 2014–2025, a range of less than four bps. Also stable

[23]The household sector includes nonprofits and residual institutions, and the corporate-equity scope includes foreign stocks held by U.S. residents.

[24]At the end of 2025, mutual funds and ETFs held $27.9 trillion in corporate equities and pension funds held $10.0 trillion, compared with $1.4 trillion held by insurance companies.

is the matched belief gap, which remains close to zero throughout: on average, respondents expect Portfolios H and L to earn about the same return.[25]

The main cross-sectional orderings documented earlier also persist over time. Among retail identities, strong Democrats have higher WTP than strong Republicans in all 12 reconstructed environments, including the two that predate the 2016 election (Figure B.7). Their average WTP is 0.73% and 0.07%, respectively, a 0.66% difference, and even the smallest difference across environments is 0.53%. The political ordering is even more pronounced for unrestricted beliefs: the average belief gap is 0.60% for strong Democrats and −0.80% for strong Republicans. The retail-institutional ordering in WTP also persists. The average WTP of the five retail identities exceeds the combined WTP of the two institutions in every information environment (Figure B.8).

The post-2020 decline in unrestricted beliefs is concentrated among the Democratic and institutional identities. From 2020 to 2023, the unrestricted belief gap falls by about 40 bps for both strong and weak Democrats and by 23 bps for the neutral identity, compared with only 6 and 2 bps for weak and strong Republicans (Figure B.7). Institutions also become more pessimistic over the same period (Figure B.8).

8.5 Checks on Mixing and Reconstruction

We conduct two checks on the historical design. The first asks whether cross-year mixing prevents a card from mapping cleanly to a single historical year. The second asks whether the reconstruction preserves the information in the model's responses by comparing reconstructed answers with answers elicited directly from single-year cards.

The first check asks the model which year from 2014 to 2025 each card most plausibly describes. We begin with 12 single-year benchmark cards, each constructed from one year's masked modules without mixing. We find that the model identifies the correct year in all 12 cases, with self-reported confidence of 100% each time (Figure 10, Panel B). Masking alone therefore leaves the year recognizable. The 288 mixed cards are constructed in the same way, except that their modules come from different years. For these cards, the model's guesses

[25]Unlike the stable reconstructed WTP, stated willingness to sacrifice dropped sharply in the retail survey of Larcker et al. (2024), although that survey asks about a one-time loss rather than an annual-return tradeoff. Between 2023 and 2024, the share of young investors willing to give up more than 10% of their retirement savings for environmental improvements fell from about 22% to about 10%.

span all 12 years, with a standard deviation of 3.0 years (Panel A). Mixing thus breaks the close correspondence between a card and a unique calendar year.

The second check asks whether mixing retains enough of the original information for the reconstructed responses to remain related to direct single-year answers without simply reproducing them. We compare the reconstructed responses with direct answers to the 12 single-year benchmark cards (Figure B.9). Across identities, the correlations between mean responses from the two methods are 1.00 for WTP, 0.90 for the matched belief gap, and 0.98 for the unrestricted belief gap. Within an identity across years, the correlations are 0.58 for WTP and 0.67 for the unrestricted belief gap.[26] The reconstruction therefore tracks the main patterns in the direct answers while remaining distinct from them.

The two checks support the intended tradeoff: cross-year mixing weakens the correspondence between a card and a single historical year, while reconstruction preserves the main patterns found in direct single-year responses.

8.6 Estimation Noise in the Reconstructed Series

We estimate each module-year effect in equation (3) from the 24 cards on which that pair appears. Part of the time variation in a reconstructed series therefore reflects estimation noise rather than a response to the information on the cards. In this subsection, we ask whether this variation exceeds what estimation noise alone would produce.

We measure this noise with the tenth module. Its assigned year labels vary and are balanced across years in the same way as the labels of the nine varying modules. Since the content never changes, any time variation estimated for it is noise. We add the module to equation (3) as a tenth set of year indicators. We then reassign its year labels at random across the 288 cards 1,000 times, with the same reassignment for all seven identities, and re-estimate the regression after each draw. The 1,000 estimated paths of the placebo module describe the time variation produced by estimation noise.

Using the ten-module extension of equation (3), we first construct an annual series for each module. For each year, we average that module's estimated coefficients across the seven identities, weighting by 2014 sector holdings. Fixed weights ensure that the series move only with the estimated coefficients, not with changes in holdings. We then add the ten module

[26]The matched belief gap remains close to zero, with little time variation, under both methods.

series and the weighted intercept to obtain the aggregate reconstructed series.[27] For each module, we regress this aggregate series on the module's own annual series over 2014–2025. The resulting R^2 measures how closely that module tracks the aggregate over time. Because the module series are correlated, the R^2 values need not sum to one.

For each placebo draw, we re-estimate the ten-module regression and calculate the placebo module's R^2 using the same procedure as for the real modules. We use the 95th percentile of the 1,000 placebo R^2 values as the cutoff. Each real module's p-value is the fraction of placebo draws with an R^2 at least as large as its own. Figure B.10 shows the distribution of R^2 values from the 1,000 placebo draws. It also reports the R^2 for the tenth module before its year labels are randomly reassigned: 25% for WTP and 3% for the unrestricted belief gap. Both are below the corresponding 95th-percentile cutoffs.

We find that several modules track the reconstructed series more closely than the placebo benchmark. The leading modules differ for tastes and beliefs, as shown in Figure 11. For WTP, policy and regulation is the only module above the placebo cutoff. For the unrestricted belief gap, economy and markets ranks first, followed by policy and regulation and then news and investor discussion. All three exceed the cutoff.

The difference is consistent with what the two questions measure. The preference question holds financial characteristics fixed and asks how much return investors would sacrifice for greenness. Policy and regulatory developments may change how investors assess the consequences of firms' environmental and social practices, and therefore how much they value greenness apart from financial returns. The unrestricted belief question instead asks about expected returns on typical real-world green and brown portfolios. These portfolios can differ in industry composition, valuations, and exposure to economic conditions. As a result, changes in growth, interest rates, and market conditions may help explain much of the time variation in the expected-return gap between them.

8.7 Relation to Real-World Series

Real-world measures of sustainable investing can reflect investors' tastes for greenness, their return beliefs, or both. Investors may buy sustainable funds because they value greenness or expect higher returns, and either motive can support fund flows, new fund launches, and

[27] Figure 9, by contrast, uses the nine-module fit and year-specific holdings weights.

higher fund fees. Attention to climate may also be associated with changes in both tastes and expected green returns. It is therefore unclear ex ante whether each real-world measure moves more closely with tastes or beliefs. We examine this question by relating our separately reconstructed WTP and unrestricted belief gap series to three annual series on sustainable investing and three on attention to climate and ESG.

After using simple correlations to describe the main patterns, we regress each of the six real-world series on both WTP and the unrestricted belief gap. These regressions show whether a real-world series is related to tastes after controlling for beliefs, and to beliefs after controlling for tastes. The two reconstructed series are only weakly correlated with each other over 2014–2025 ($\rho = 0.15$), making this comparison informative. Table 8 reports regressions with variables standardized within each sample. The reported t-statistics use OLS standard errors, with p-values from the t distribution with $n - 3$ degrees of freedom. Table B.7 gives the sources, definitions, and sample periods of the real-world series.

We first examine three measures of sustainable investing: net flows into U.S. sustainable funds, next-year sustainable-fund launches, and the ESG fee estimated by Baker et al. (2024). Flows measure net investment in sustainable funds, while launches measure the creation of new funds. We match WTP in year t with launches in year $t+1$, allowing for the time required to bring a new fund to market. The ESG fee is the additional annual fee that investors are willing to pay for an index fund with an ESG mandate, estimated from index-fund flows. It is the measure closest in concept to stated WTP.

Figure 12 plots the three investing series along with the reconstructed WTP. All three series are positively correlated with WTP. The correlations are 0.79 for flows and 0.77 for next-year launches over 2014–2024 ($p < 0.01$ for both), and 0.90 for the ESG fee over 2019–2023 ($p < 0.05$). These relationships still hold when we control for the unrestricted belief gap. In multiple regressions, the standardized WTP coefficients range from 0.70 to 0.78 and are significant at the 1% level for flows, 5% for launches, and 10% for the fee. None of the belief coefficients is significant (Table 8).[28]

We next examine three measures of attention to climate and ESG. The Media Climate Change Concerns index of Ardia et al. (2023) tracks climate-related news coverage, U.S.

[28]The ESG fee requires some caution. Its strong correlation with WTP rests on only five annual observations. The fee can also reflect return beliefs (Baker et al., 2024), and its simple correlation with the belief gap is 0.76 ($p = 0.14$). Moreover, the fee ranges from -33 to 17 bps over 2019–2023, much wider than the 3.4-bps range of reconstructed WTP over 2014–2025.

Google searches for ESG measure public search interest, and the climate-change exposure of earnings calls from Sautner et al. (2023) measures how much climate change is discussed by managers and analysts. We compare these series with the unrestricted belief gap in equation (2). A more negative gap means that green portfolios are expected to underperform brown portfolios by more.

Figure 13 plots the three attention series along with the reconstructed unrestricted belief gap. The gap is negatively correlated with all three series. The correlations are -0.85 for media concern and -0.90 for Google search interest over 2014–2025, and -0.84 for earnings-call exposure over 2014–2024, all with $p < 0.01$. The relationships remain significant after controlling for WTP. In multiple regressions, the standardized belief coefficients range from -0.88 to -0.96, all significant at the 1% level (Table 8). WTP is not significantly related to media concern or search interest. It is significantly positively related to earnings-call exposure, though the R-squared from that simple regression is less than 0.04. One interpretation is that managers discuss climate more when investors value greenness more, consistent with firms responding to investor demand in other settings (Baker and Wurgler, 2004).

We also examine whether these relationships are sensitive to general market sentiment. When we add sentiment controls, the WTP coefficients for flows and launches and the belief coefficients for the three attention series keep their signs and remain statistically significant in nearly all specifications (Tables B.8 and B.9).[29]

One limitation of these comparisons is the short time series, which is necessitated by data availability. In such short series, correlations can be substantial even between unrelated variables, so we view these comparisons as descriptive. Another limitation is that some of the real-world information also appears on the cards we use to reconstruct tastes and beliefs. The sustainable-investment market module reports U.S. sustainable-fund flows for seven of the 12 years. However, its annual effects have the lowest R^2 with reconstructed WTP among the nine year-varying modules, at just 0.1% (Figure 11, Panel A).

[29]The controls are the San Francisco Fed news sentiment index (Shapiro et al., 2022), the University of Michigan consumer sentiment index, and the VIX, added separately and jointly. The fee sample is too short for these regressions. The WTP coefficients for flows and launches and both the WTP and belief coefficients for earnings-call exposure remain significant at the 5% level or better throughout. The belief coefficients for media concern and search interest remain negative and generally significant, although the former becomes insignificant when all three controls are included.

Sustainable-investing theory offers one explanation for the negative relation between climate attention and our reconstructed belief gap. In Pastor et al. (2021), stronger demand for green assets raises their prices and lowers their expected returns relative to brown assets. An unexpected rise in demand can nevertheless produce positive realized green returns as prices adjust. Green stocks have indeed outperformed brown stocks when climate concerns rose unexpectedly (Choi et al., 2020). Pastor et al. (2022) and Ardia et al. (2023) document both this outperformance and lower expected green returns. If attention raises green demand, the decline in our belief gap thus matches the theory's prediction for expected returns. This demand need not come from stronger tastes; it could reflect climate-risk hedging or outside pressure of the kind our institutional identities cite (Section 6).

Return extrapolation would predict a different pattern. Investors who extrapolate from the green outperformance associated with increases in climate attention would expect green portfolios to continue outperforming. Prior research documents return extrapolation in human expectations of aggregate-market returns (Greenwood and Shleifer, 2014) and individual-stock returns (Da et al., 2021). Our negative attention–belief relationship instead has the direction predicted by equilibrium pricing.

These findings also help us compare the roles of tastes and beliefs in sustainable-fund demand. In Pastor et al. (2021), stronger tastes for greenness can coincide with lower expected green returns, so fund flows alone cannot distinguish changes in tastes from changes in beliefs. Our multiple regressions show that flows and subsequent launches are more closely related to reconstructed tastes than to reconstructed beliefs. This pattern is consistent with tastes playing an important role in the time variation of sustainable-fund demand. The evidence is limited, however: both regressions rely on only 11 annual observations, so the insignificant belief coefficients do not rule out a role for return beliefs.

9 Conclusion

We use LLMs as synthetic survey respondents to elicit tastes for greenness, return beliefs, portfolio allocations, and investment motives separately from the same retail and institutional investor identities. We show that a greener portfolio tilt does not necessarily indicate a stronger taste for greenness. Compared with retail identities, institutional identities state weaker green tastes yet allocate more of their equity to the greener of two otherwise identical

portfolios. Beliefs do not explain this tilt. Instead, institutions attribute roughly half of their greener-portfolio holding to outside pressure, whereas retail mostly cites its own choice. These differences hold across all four full-grid models, and the stronger institutional tilt remains after we control for tastes and matched-portfolio beliefs. Inferring the distribution of green tastes from portfolio holdings can therefore be misleading, especially for institutions.

Tastes and beliefs also behave differently across the three ESG dimensions and over time. Governance is closely related to beliefs, with respondents expecting high-governance portfolios to outperform, but not to tastes. In the reconstructed 2014–2025 information environments, beliefs about typical real-world green portfolios become more pessimistic after 2020, whereas tastes vary little. Tastes comove with sustainable-fund flows, subsequent fund launches, and the fee investors are willing to pay for ESG funds, while beliefs move inversely with attention to climate and ESG. This pattern is consistent with equilibrium pricing: if greater attention raises demand for green assets, their prices rise and expected returns fall.

Green tastes are closely related to investor characteristics. Among retail identities, political affiliation is by far the strongest predictor of WTP. WTP also rises with income but falls with financial literacy. Among institutional identities, type and geography matter most, with the highest WTP among endowments, foundations, and European institutions. Recessions lower WTP for retail but not for institutions.

Our design has several advantages over a human survey. The complete factorial grid makes investor characteristics orthogonal by construction, allowing us to vary one characteristic at a time and to compare their relative importance without the correlations present in observational samples. The same identity can answer in a recession, in a boom, or in a reconstructed historical information environment. One set of questions goes to both institutional and retail identities. Attrition and fatigue are absent: no identity skips a question or a condition. Separate model calls allow us to compare tastes, beliefs, allocations, and motives without one answer contaminating another.

The main limitation is that these are model-generated responses rather than observations of human investors. Our validation exercises reproduce key signs and cross-sectional orderings of retail beliefs in human survey data, but not the dispersion of response levels. We therefore place greater weight on cross-sectional orderings and comparative statics that recur across models. Synthetic surveys do not replace human surveys, but they can complement them

by enabling controlled comparisons across characteristics, populations, and counterfactual environments that are difficult to obtain in conventional samples.

More broadly, our framework can be extended to other non-pecuniary attributes, investor types, and historical settings for which the relevant information environment can be assembled.

References

Anthis, Jacy Reese, Ryan Liu, Sean M. Richardson, Austin C. Kozlowski, Bernard Koch, Erik Brynjolfsson, James Evans, and Michael S. Bernstein, 2025, Position: LLM social simulations are a promising research method, in *Proceedings of the 42nd International Conference on Machine Learning*, Proceedings of Machine Learning Research 267, 81005–81034.

Ardia, David, Keven Bluteau, Kris Boudt, and Koen Inghelbrecht, 2023, Climate change concerns and the performance of green vs. brown stocks, *Management Science* 69, 7607–7632.

Argyle, Lisa P., Ethan C. Busby, Nancy Fulda, Joshua R. Gubler, Christopher Rytting, and David Wingate, 2023, Out of one, many: Using language models to simulate human samples, *Political Analysis* 31, 337–351.

Aron-Dine, Shifrah, Johannes Beutel, Monika Piazzesi, and Martin Schneider, 2025, Household climate finance: Theory and survey data on safe and risky green assets, Working paper, National Bureau of Economic Research, NBER Working Paper 32615, revised March 2025.

Ashokkumar, Ashwini, Luke Hewitt, Isaias Ghezae, and Robb Willer, 2026, Large language models can predict the results of social science experiments, *Nature* 656, 115–122.

Baker, Malcolm, Mark L. Egan, and Suproteem K. Sarkar, 2024, Demand for ESG, Working paper, National Bureau of Economic Research, NBER Working Paper 30708, revised December 2024.

Baker, Malcolm, and Jeffrey Wurgler, 2004, A catering theory of dividends, *Journal of Finance* 59, 1125–1165.

Bansal, Ravi, Di Wu, and Amir Yaron, 2022, Socially responsible investing in good and bad times, *Review of Financial Studies* 35, 2067–2099.

Barber, Brad M., Adair Morse, and Ayako Yasuda, 2021, Impact investing, *Journal of Financial Economics* 139, 162–185.

Bauer, Rob, Tobias Ruof, and Paul Smeets, 2021, Get real! Individuals prefer more sustainable investments, *Review of Financial Studies* 34, 3976–4043.

Bauer, Rob, Katrin Gödker, Paul Smeets, and Florian Zimmermann, 2026, Mental models in financial markets: How do experts reason about the pricing of climate risk?, Working paper, Maastricht University, Bocconi University, University of Amsterdam, and University of Bonn, July 2026 version.

Bhagwat, Vineet, J. Anthony Cookson, Chukwuma Dim, and Marina Niessner, 2026, The market's mirror: Revealing investor disagreement with LLMs, Working paper, SSRN Working Paper 5375473, July 2026 version.

Bisbee, James, Joshua D. Clinton, Cassy Dorff, Brenton Kenkel, and Jennifer M. Larson, 2024, Synthetic replacements for human survey data? The perils of large language models, *Political Analysis* 32, 401–416.

Board of Governors of the Federal Reserve System, 2023, 2022 Survey of Consumer Finances: Public-use data and documentation, dataset and technical documentation, `https://www.federalreserve.gov/econres/scfindex.htm`.

Board of Governors of the Federal Reserve System, 2026, Financial Accounts of the United States, Z.1, June 11, 2026 release, `https://www.federalreserve.gov/releases/z1/20260611/z1.pdf`.

Bonnefon, Jean-François, Augustin Landier, Parinitha Sastry, and David Thesmar, 2025, The moral preferences of investors: Experimental evidence, *Journal of Financial Economics* 163, 103955.

Brand, James, Ayelet Israeli, and Donald Ngwe, 2026, Using LLMs for market research, Working paper, Harvard Business School, HBS Working Paper 23-062, April 2026 version.

Brodback, Daniel, Nadja Guenster, Sébastien Pouget, and Ruichen Wang, 2026, Investor valuation for socially responsible assets: A willingness to pay experiment, *Management Science* 72(7), 5544–5560.

Bybee, J. Leland, 2025, The ghost in the machine: Generating beliefs with large language models, Working paper, University of Chicago Booth School of Business, February 2025 version.

CFA Institute, 2022, Enhancing investors' trust: 2022 CFA Institute Investor Trust Study, Survey report, CFA Institute, `https://www.cfainstitute.org/sites/default/files/-/media/documents/article/Enhancing-Investors-Trust-Report_2022_Online.pdf`.

Choi, Darwin, Zhenyu Gao, and Wenxi Jiang, 2020, Attention to global warming, *Review of Financial Studies* 33, 1112–1145.

Da, Zhi, Xing Huang, and Lawrence J. Jin, 2021, Extrapolative beliefs in the cross-section: What can we learn from the crowds?, *Journal of Financial Economics* 140, 175–196.

Dahlquist, Magnus, and Markus Ibert, 2024, Equity return expectations and portfolios: Evidence from large asset managers, *Review of Financial Studies* 37, 1887–1928.

Döttling, Robin, and Sehoon Kim, 2024, Sustainability preferences under stress: Evidence from COVID-19, *Journal of Financial and Quantitative Analysis* 59, 435–473.

Edmans, Alex, Tom Gosling, and Dirk Jenter, 2026, Sustainable investing in practice: Objectives, beliefs, and limits to impact, Working paper, SSRN Working Paper 4963062, January 30, 2026 version.

Engelberg, Joseph, Asaf Manela, William Mullins, and Luka Vulicevic, 2026, Entity neutering, Working paper, SSRN Working Paper 5182756, June 26, 2026 version.

Engle, Robert F., Stefano Giglio, Bryan Kelly, Heebum Lee, and Johannes Stroebel, 2020, Hedging climate change news, *Review of Financial Studies* 33, 1184–1216.

Eskildsen, Marc, Markus Ibert, Theis Ingerslev Jensen, and Lasse Heje Pedersen, 2026, Realized returns to green investing: A global replication, Working paper, SSRN Working Paper 6527562, April 6, 2026 version.

Fedyk, Anastassia, Ali Kakhbod, Peiyao Li, and Ulrike Malmendier, 2026, AI and perception biases in investments: An experimental study, *Journal of Financial Economics* 185, 104350.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus, 2021, Five facts about beliefs and portfolios, *American Economic Review* 111, 1481–1522.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, Zhenhao Tan, Stephen Utkus, and Xiao Xu, 2025, Four facts about ESG beliefs and investor portfolios, *Journal of Financial Economics* 164, 103984.

Gompers, Paul, Joy Ishii, and Andrew Metrick, 2003, Corporate governance and equity prices, *Quarterly Journal of Economics* 118, 107–156.

Graham, John R., Campbell R. Harvey, and Manish Jha, 2026, CFOs meet LLMs, Working paper, arXiv:2606.13812, September 2026 version.

Greenwood, Robin, and Andrei Shleifer, 2014, Expectations of returns and expected returns, *Review of Financial Studies* 27, 714–746.

Groves, Robert M., Stanley Presser, and Sarah Dipko, 2004, The role of topic interest in survey participation decisions, *Public Opinion Quarterly* 68, 2–31.

Hansen, Anne Lundgaard, John J. Horton, Sophia Kazinnik, Daniela Puzzello, and Ali Zarifhonarvar, 2026, Simulating the Survey of Professional Forecasters, Working paper, SSRN Working Paper 5066286, June 2026 version.

Hartzmark, Samuel M., and Abigail B. Sussman, 2019, Do investors value sustainability? A natural experiment examining ranking and fund flows, *Journal of Finance* 74, 2789–2837.

He, Songrun, Linying Lv, Asaf Manela, and Jimmy Wu, 2025, Chronologically consistent large language models, *Journal of Financial Economics*, forthcoming.

Heeb, Florian, Julian F. Kölbel, Falko Paetzold, and Stefan Zeisberger, 2023, Do investors care about impact?, *Review of Financial Studies* 36, 1737–1787.

Hong, Harrison, and Leonard Kostovetsky, 2012, Red and blue investing: Values and finance, *Journal of Financial Economics* 103, 1–19.

Horton, John J., Apostolos Filippas, and Benjamin S. Manning, 2026, Large language models as simulated economic agents: What can we learn from homo silicus?, Working paper,

National Bureau of Economic Research, NBER Working Paper 31122, revised February 2026.

Humphrey, Jacquelyn, Shimon Kogan, Jacob S. Sagi, and Laura T. Starks, 2023, The asymmetry in responsible investing preferences and beliefs, Working paper, SSRN Working Paper 3583862, September 2023 version.

Kelly, Bryan T., Semyon Malamud, Johannes Schwab, and Teng Andrea Xu, 2026, Scaling point-in-time language models, Working paper, National Bureau of Economic Research, NBER Working Paper 35247, May 2026.

Krueger, Philipp, Zacharias Sautner, and Laura T. Starks, 2020, The importance of climate risks for institutional investors, *Review of Financial Studies* 33, 1067–1111.

Larcker, David F., Amit Seru, and Brian Tayan, 2024, 2024 survey of investors, retirement savings, and ESG, Survey report (abbreviated version), Stanford Graduate School of Business, Hoover Institution Working Group on Corporate Governance, and Rock Center for Corporate Governance, `https://www.hoover.org/sites/default/files/2024-10/cgri-survey-2024-investors-retirement-savings-esg.pdf`.

Levy, Bradford, 2026, Caution ahead: Numerical reasoning and look-ahead bias in AI models, *Journal of Accounting Research* 64, 1139–1188.

Li, Tong, and Luping Yu, 2025, ESG beliefs of large language models: Evidence and impact, Working paper, arXiv:2601.00836v1, December 2025 version.

Lopez-Lira, Alejandro, Yuehua Tang, and Mingyin Zhu, 2026, The memorization problem: Can we trust LLMs' economic forecasts?, Working paper, SSRN Working Paper 5217505, April 2, 2026 version.

Lusardi, Annamaria, and Olivia S. Mitchell, 2014, The economic importance of financial literacy: Theory and evidence, *Journal of Economic Literature* 52, 5–44.

Manski, Charles F., 2004, Measuring expectations, *Econometrica* 72, 1329–1376.

MSCI, 2026, ESG Ratings Methodology, March 2026, MSCI Sustainability and Climate Methodologies.

Nagel, Stefan, Claire Tseng, and Dacheng Xiu, 2026, Completions, not convictions: Can LLM probabilities proxy for human beliefs?, Working paper, University of Chicago, August 11, 2026 version.

Pastor, Lubos, Robert F. Stambaugh, and Lucian A. Taylor, 2021, Sustainable investing in equilibrium, *Journal of Financial Economics* 142, 550–571.

Pastor, Lubos, Robert F. Stambaugh, and Lucian A. Taylor, 2022, Dissecting green returns, *Journal of Financial Economics* 146, 403–424.

Pastor, Lubos, Robert F. Stambaugh, and Lucian A. Taylor, 2025a, Green tilts, *Journal of Financial Economics* 174, 104173.

Pastor, Lubos, Robert F. Stambaugh, and Lucian A. Taylor, 2025b, Sustainable investing, *Annual Review of Financial Economics* 17, 321–341.

Pastor, Lubos, and M. Blair Vorsatz, 2020, Mutual fund performance and flows during the COVID-19 crisis, *Review of Asset Pricing Studies* 10, 791–833.

Pedersen, Lasse Heje, Shaun Fitzgibbons, and Lukasz Pomorski, 2021, Responsible investing: The ESG-efficient frontier, *Journal of Financial Economics* 142, 572–597.

Riedl, Arno, and Paul Smeets, 2017, Why do investors hold socially responsible mutual funds?, *Journal of Finance* 72, 2505–2550.

Santurkar, Shibani, Esin Durmus, Faisal Ladhak, Cinoo Lee, Percy Liang, and Tatsunori Hashimoto, 2023, Whose opinions do language models reflect?, in *Proceedings of the 40th International Conference on Machine Learning*, Proceedings of Machine Learning Research 202, 29971–30004.

Sarkar, Suproteem K., and Keyon Vafa, 2024, Lookahead bias in pretrained language models, Working paper, SSRN Working Paper 4754678, October 18, 2024 version.

Sautner, Zacharias, Laurence van Lent, Grigory Vilkov, and Ruishen Zhang, 2023, Firm-level climate change exposure, *Journal of Finance* 78, 1449–1498.

Shapiro, Adam Hale, Moritz Sudhof, and Daniel J. Wilson, 2022, Measuring news sentiment, *Journal of Econometrics* 228, 221–243.

Stantcheva, Stefanie, 2023, How to run surveys: A guide to creating your own identifying variation and revealing the invisible, *Annual Review of Economics* 15, 205–234.

Starks, Laura T., 2023, Presidential address: Sustainable finance and ESG issues—value versus values, *Journal of Finance* 78, 1837–1872.

Wang, Angelina, Jamie Morgenstern, and John P. Dickerson, 2025, Large language models that replace human participants can harmfully misportray and flatten identity groups, *Nature Machine Intelligence* 7, 400–411.

Wu, Jing Cynthia, Jin Xi, and Shihan Xie, 2025, LLM survey framework: Coverage, reasoning, dynamics, identification, Working paper, National Bureau of Economic Research, NBER Working Paper 34308, October 2025.

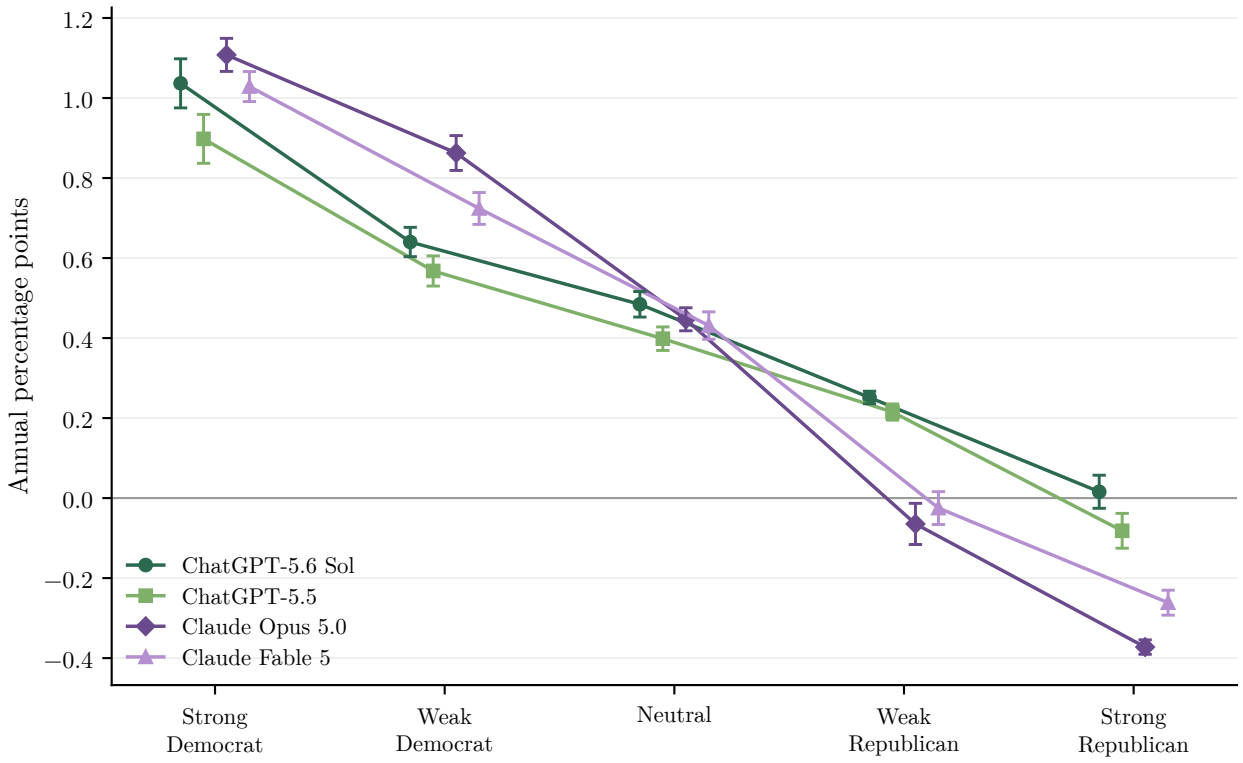


Figure 1. Retail WTP and political affiliation. This figure shows mean willingness to pay (WTP) for greenness by political affiliation in the four models that answer the full grid, in annual percentage points. WTP is the average of the annual return given up for Portfolio H and the extra annual return required for Portfolio L over ten years. The portfolios are otherwise identical and score at the 80th (H) and 20th (L) percentiles of the same greenness measure. The sample is the 480 baseline retail identities of each model, with aggregate ESG scores and a normal economy. Each point averages the 96 baseline retail cells at that political level in the indicated model. For ChatGPT-5.6 Sol, the three repeats are averaged within each cell first. Error bars are 95% confidence intervals, calculated as 1.96 standard errors across the 96 cells. These intervals summarize cross-cell variation in synthetic answers, rather than sampling uncertainty about a human population. The four models are offset horizontally for legibility.

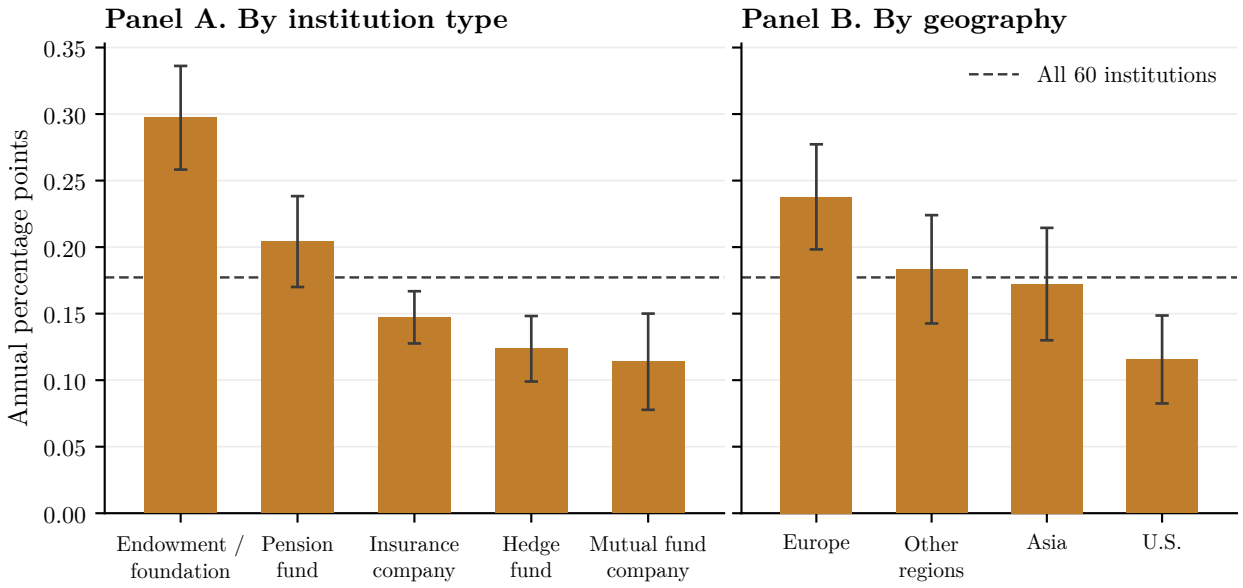


Figure 2. Institutional WTP by type and geography. This figure plots institutional willingness to pay (WTP) by institution type and geography in ChatGPT-5.6 Sol, in annual percentage points. WTP is the average of the annual return given up for Portfolio H and the extra annual return required for Portfolio L over ten years. The portfolios are otherwise identical and score at the 80th (H) and 20th (L) percentiles of the same greenness measure. The baseline uses aggregate ESG scores and a normal economy. Panel A averages the 12 cells of each institution type, and Panel B averages the 15 cells of each geography. The three repeats are averaged within each cell. The dashed line is the mean across all 60 institutions. Error bars are 95% confidence intervals, calculated as 1.96 standard errors across cells.

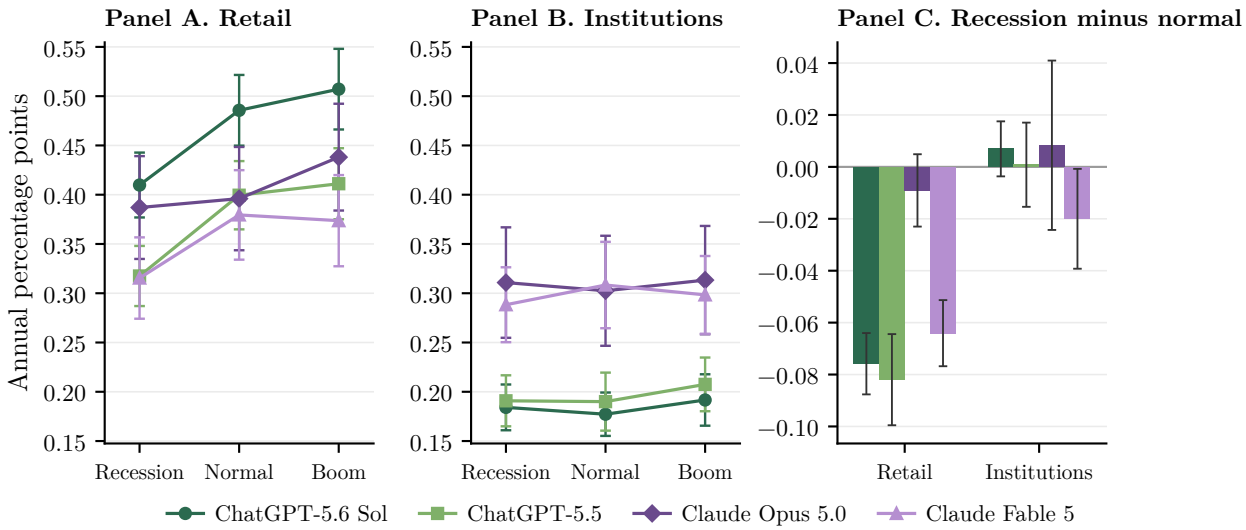


Figure 3. Retail and institutional WTP by economy state. This figure plots willingness to pay (WTP) by economy state in the four models that answer the full grid, in annual percentage points. WTP is the average of the annual return given up for Portfolio H and the extra annual return required for Portfolio L over ten years. The portfolios are otherwise identical and score at the 80th (H) and 20th (L) percentiles of the same greenness measure. Panels A and B show the 480 retail and 60 institutional identities, respectively, on the same vertical scale. Each point averages the 480 retail cells (Panel A) or 60 institutional cells (Panel B) in that economy state for the indicated model. For ChatGPT-5.6 Sol, the three repeats are averaged within each cell first. Error bars are 95% confidence intervals, calculated as 1.96 standard errors across cells. Panel C reports the recession-minus-normal difference for retail identities on the left and institutions on the right. Differences are computed within identity from cell means and then averaged; the corresponding intervals use 1.96 standard errors of the paired differences.

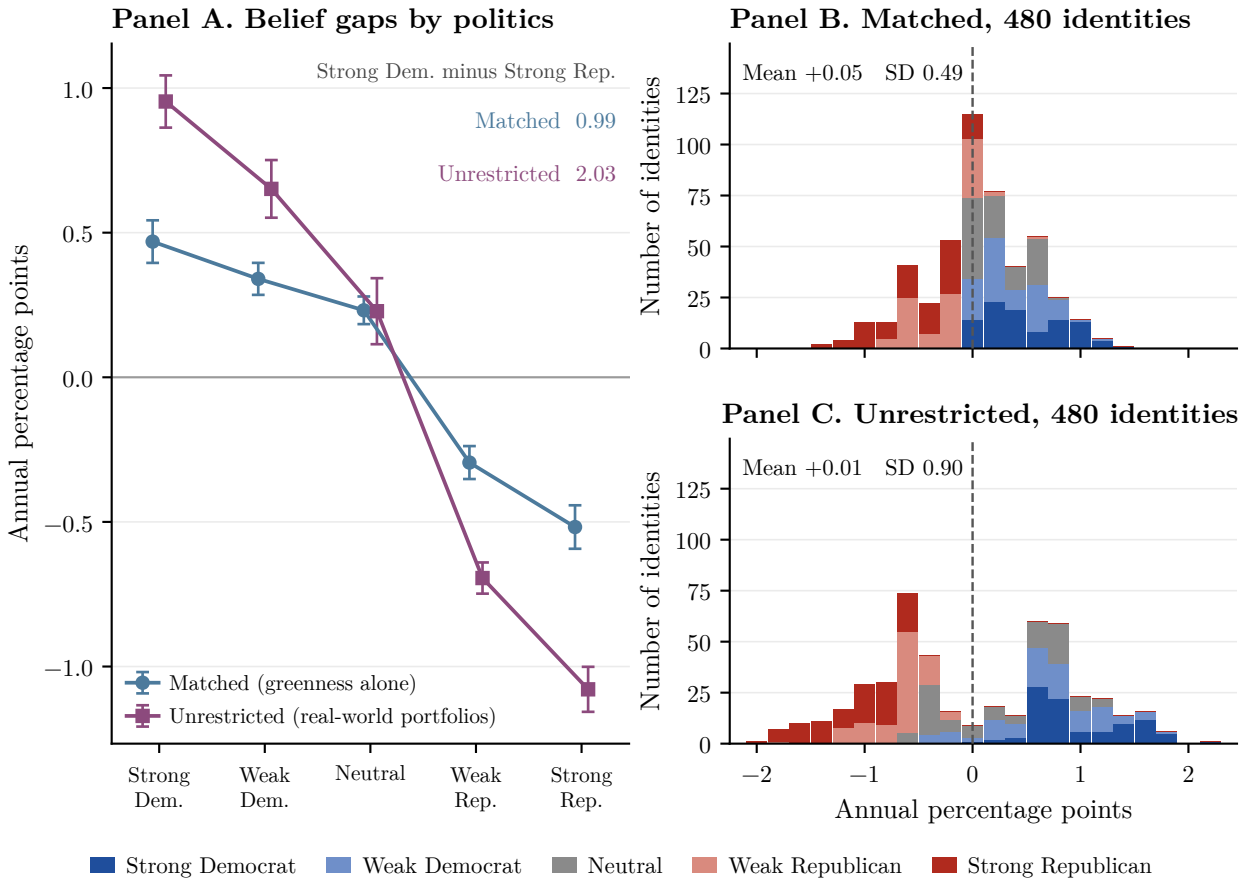


Figure 4. Matched and unrestricted belief gaps of retail identities. This figure compares matched and unrestricted belief gaps across the 480 baseline retail identities in ChatGPT-5.6 Sol, in annual percentage points. Both belief gaps are expected annual returns over ten years on portfolios at the 80th score percentile minus those at the 20th percentile. The matched question holds all other portfolio characteristics fixed; the unrestricted question refers to typical real-world portfolios at those percentiles. The baseline uses aggregate ESG scores and a normal economy. Panel A reports the mean matched gap (circles) and unrestricted gap (squares) by political affiliation. Each point averages the 96 baseline cell means at that political level. Error bars are 95% confidence intervals, calculated as 1.96 standard errors across the 96 cells; the printed values are the strong-Democrat-minus-strong-Republican spreads. Panels B and C show the matched and unrestricted distributions, respectively, after averaging the three repeats within each cell. Bars are stacked by political affiliation on a common horizontal scale, with a bin width of 0.2 percentage points. The printed values are the mean and standard deviation across identities, and the dashed line marks zero.

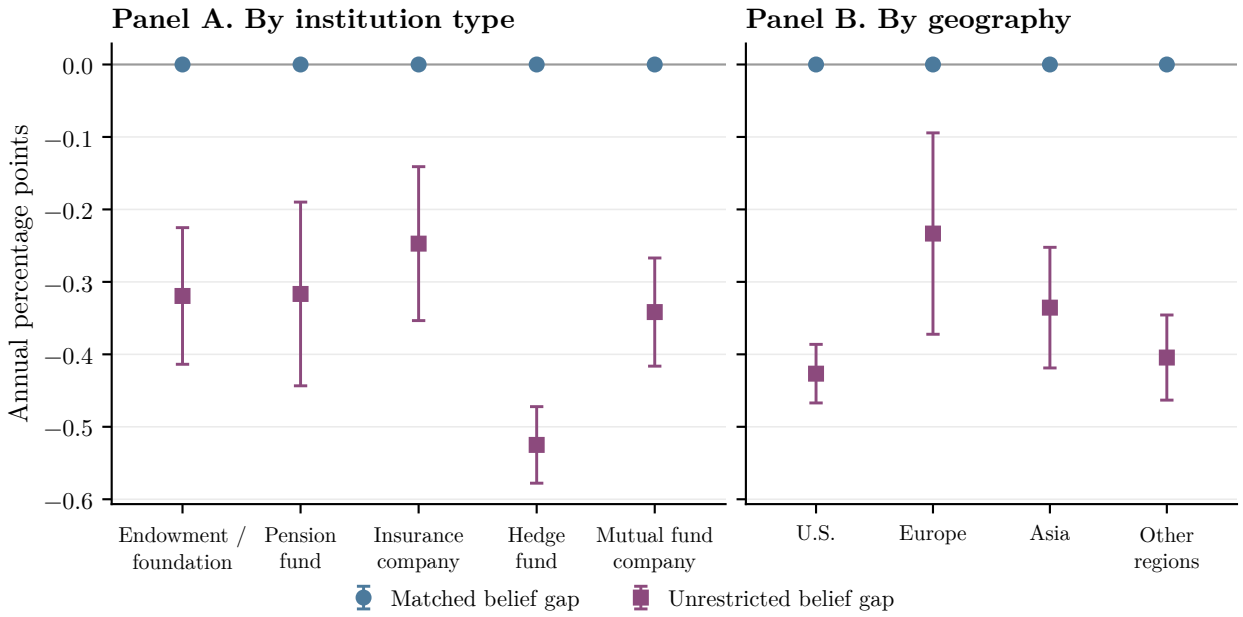


Figure 5. Institutional belief gaps by type and geography. This figure plots institutional matched and unrestricted belief gaps by institution type and geography in ChatGPT-5.6 Sol, in annual percentage points. Both belief gaps are expected annual returns over ten years on portfolios at the 80th score percentile minus those at the 20th percentile. The matched question holds all other portfolio characteristics fixed; the unrestricted question refers to typical real-world portfolios at those percentiles. The baseline uses aggregate ESG scores and a normal economy. Panel A averages the 12 baseline cells of each institution type, and Panel B averages the 15 baseline cells of each geography. The three repeats are averaged within each cell. Circles denote matched belief gaps, and squares denote unrestricted belief gaps. Error bars are 95% confidence intervals, calculated as 1.96 standard errors across cells.

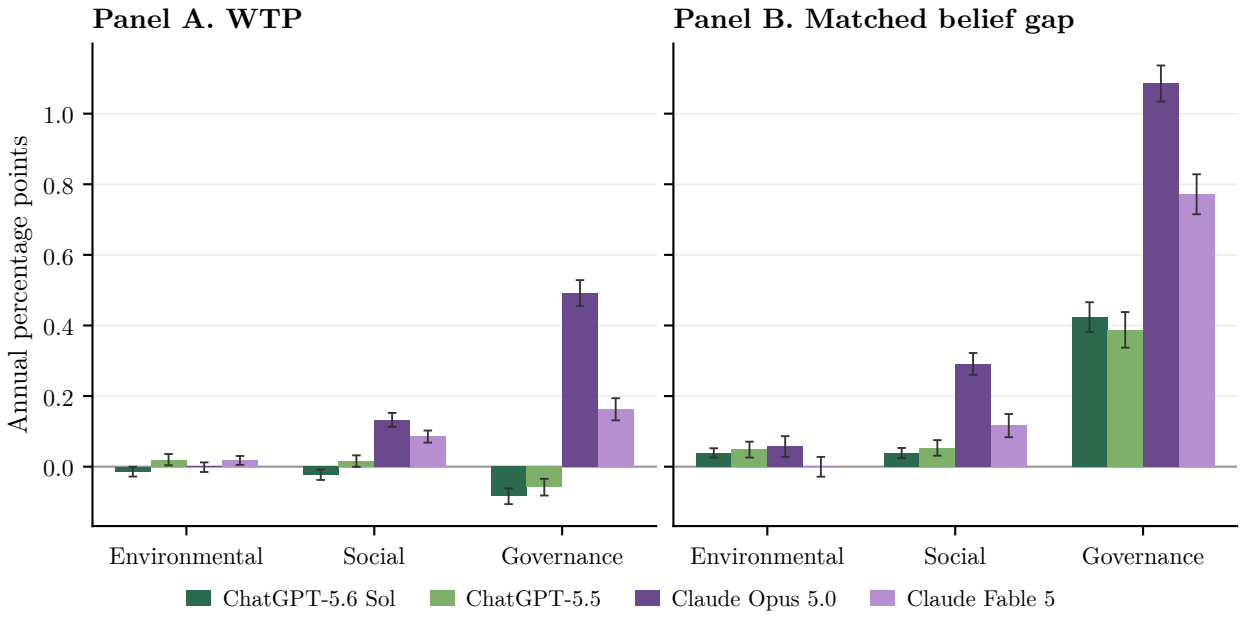


Figure 6. Changes in tastes and beliefs across ESG dimensions. This figure plots within-identity changes in willingness to pay (WTP) and in the matched belief gap when the score measures only environmental, social, or governance characteristics instead of aggregate ESG. The four models answer the full grid, and each bar pools 540 paired observations (480 retail and 60 institutional identities). Panel A reports WTP, the average of the annual return given up for Portfolio H and the extra annual return required for Portfolio L over ten years. Panel B reports the matched belief gap, the expected annual return over ten years on Portfolio H minus that on Portfolio L. Both are in annual percentage points. Portfolios H and L are otherwise identical and score at the 80th and 20th percentiles of the same greenness measure. The economy is normal. Each difference subtracts the same identity's aggregate-ESG baseline response, so positive values indicate a larger WTP or a larger gap under the individual dimension. For ChatGPT-5.6 Sol, the three repeats are averaged within each cell. The two panels share one vertical scale. Error bars are 95% confidence intervals, calculated as 1.96 standard errors of the paired differences. Figure B.2 reports the changes separately for retail and institutional identities.

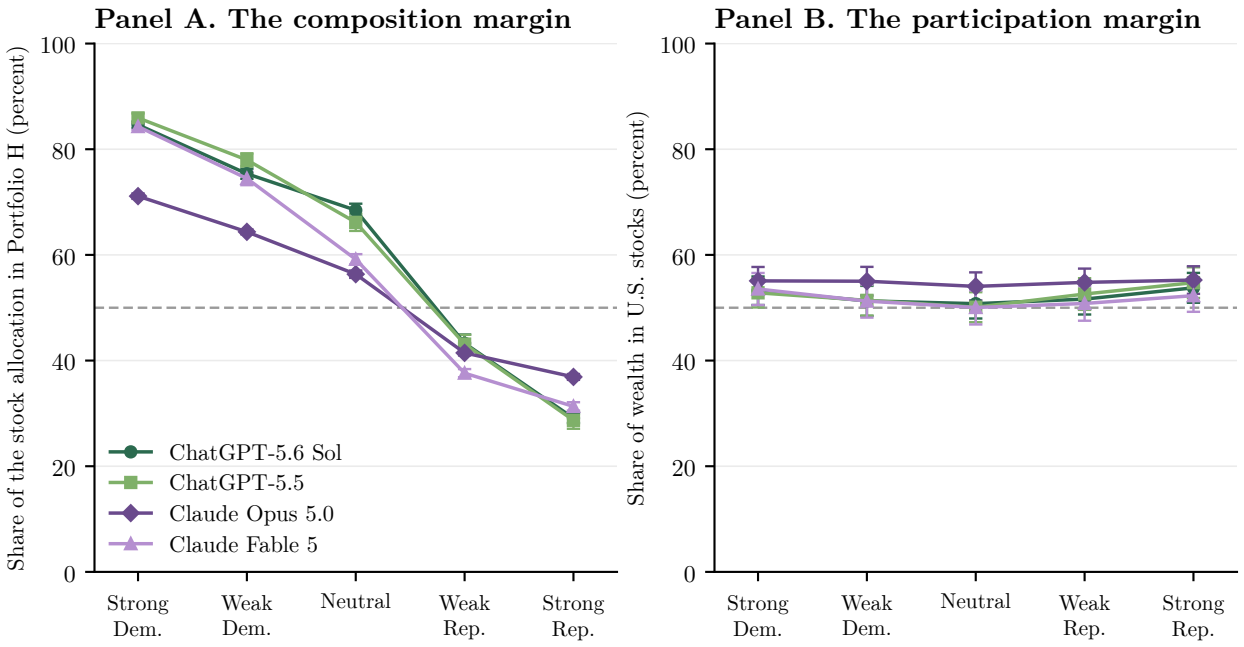


Figure 7. Portfolio allocations and political affiliation. This figure plots stated portfolio allocations by political affiliation in the four models that answer the full grid, in percent. Portfolios H and L are otherwise identical and score at the 80th and 20th percentiles of the same greenness measure. The baseline uses aggregate ESG scores and a normal economy. Each point averages the 96 baseline retail cells at that political level, after averaging the three repeats within each cell for ChatGPT-5.6 Sol. Panel A shows the share of the stock allocation assigned to Portfolio H. The question offers H and L as the only equity choices, and the dashed line at 50 denotes an equal split. Panel B shows the share of financial wealth assigned to U.S. stocks on the same vertical scale; its line at 50 is a scale reference. Error bars are 95% confidence intervals, calculated as 1.96 standard errors across cells.

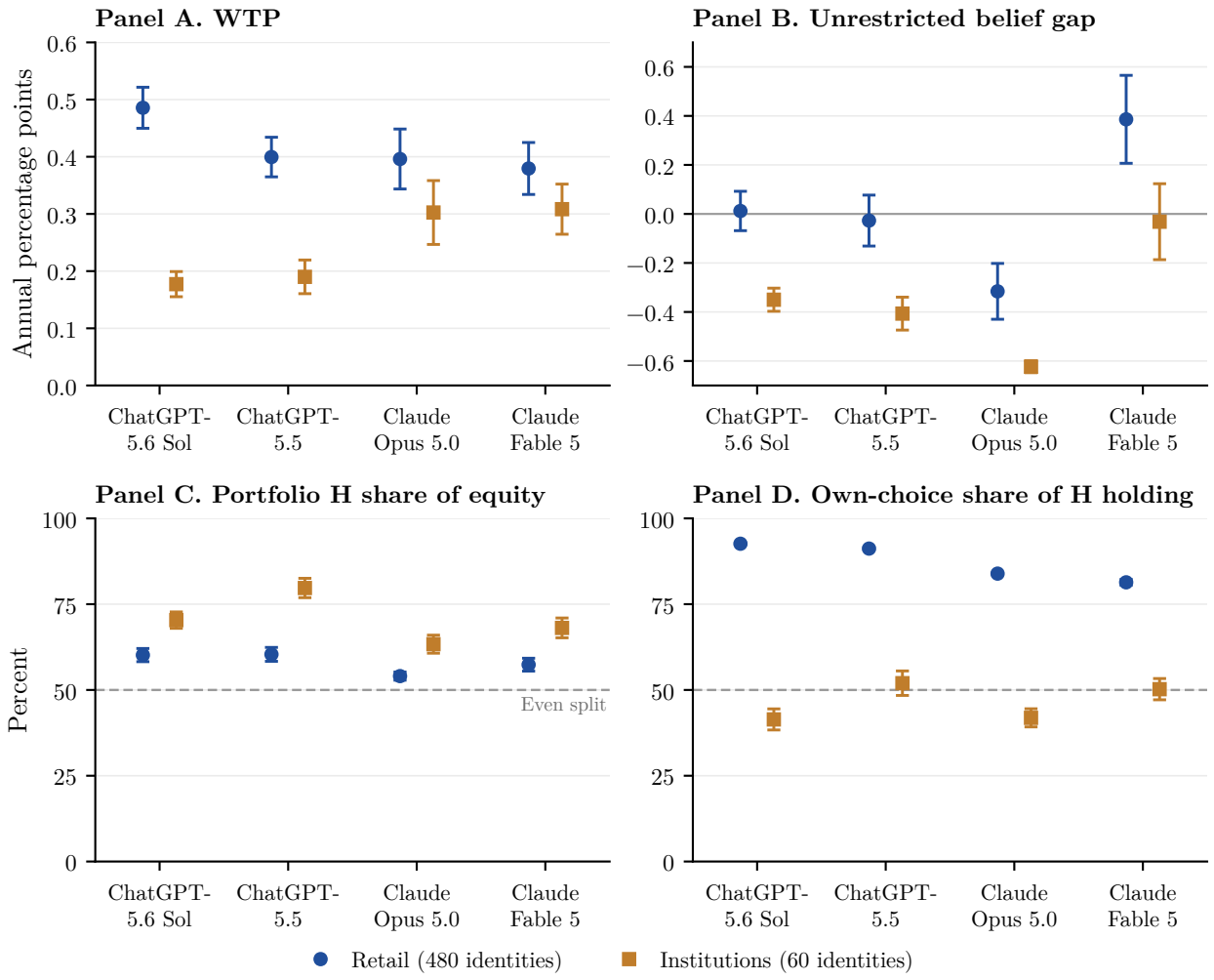


Figure 8. Tastes, beliefs, allocations, and motives of retail and institutional identities. This figure compares retail and institutional identities in the four models that answer the full grid. Panel A plots willingness to pay (WTP), the average of the annual return given up for Portfolio H and the extra annual return required for Portfolio L over ten years. Panel B plots the unrestricted belief gap, the expected annual return over ten years on typical real-world portfolios at the 80th score percentile minus that at the 20th. Both are in annual percentage points. Panel C plots the percentage of the stated equity allocation assigned to Portfolio H, and Panel D the percentage of the stated Portfolio H holding attributed to own choice rather than outside pressure. In Panels A, C, and D, H and L are otherwise identical portfolios at the 80th and 20th percentiles of the same greenness measure. The dashed lines at 50 in Panels C and D mark an even split. The sample is the 480 retail and 60 institutional identities, with aggregate ESG scores and a normal economy. ChatGPT-5.6 Sol's three repeats are averaged within each cell. Error bars are 95% confidence intervals, calculated as 1.96 standard errors across cells, and some are narrower than the markers.

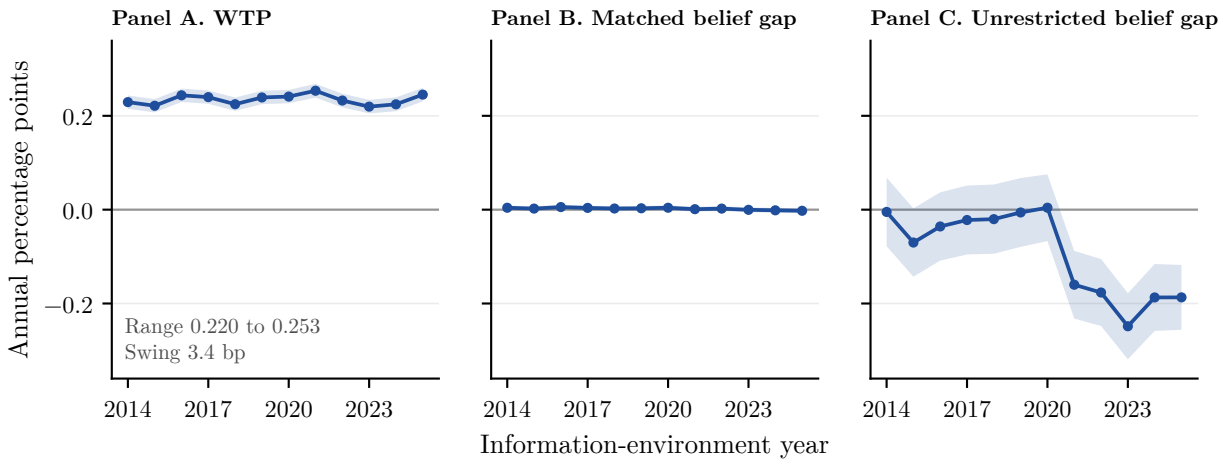


Figure 9. Reconstructed tastes and beliefs, 2014–2025. This figure shows the reconstructed series for the 2014–2025 information environments in ChatGPT-5.6 Sol, weighted by three sector holdings proxies. Panels A, B, and C report willingness to pay (WTP), the matched belief gap, and the unrestricted belief gap, respectively, in annual percentage points on a common vertical scale. WTP averages the annual return given up for Portfolio H and the extra return required for Portfolio L; belief gaps are the expected return on H minus that on L. Both refer to annual returns over ten years. H and L score at the 80th and 20th percentiles of the ESG measure; they are otherwise identical in the preference and matched-belief questions, whereas unrestricted beliefs refer to typical real-world portfolios. Panel A prints the range, in percentage points, and the swing, in bps. Each series combines the mean of five retail political identities, a mutual fund company, and a pension fund using annual corporate-equity holdings of the household, mutual fund and ETF, and pension sectors from the Federal Reserve's Z.1 accounts (Board of Governors, 2026), normalized over these sectors. The household sector includes nonprofits and residual institutions; the equity scope includes foreign stocks held by U.S. residents. Shaded bands are 95% confidence intervals, calculated as 1.96 reconstruction standard errors. The weighted errors incorporate the estimated error correlation across identities, each of which answers the same 288 cards.

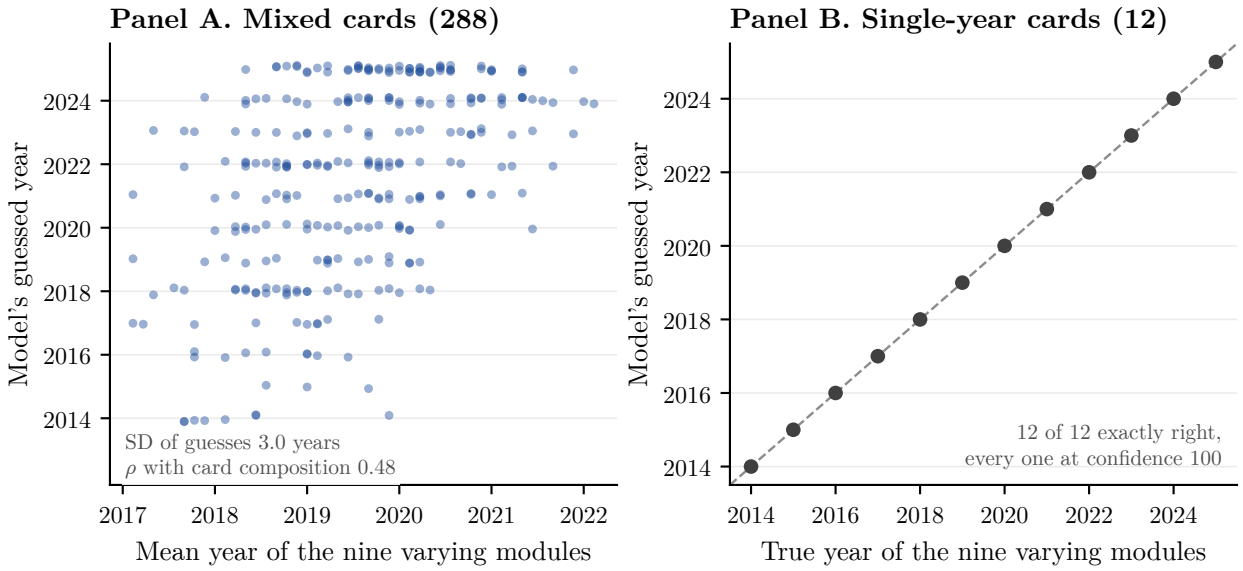


Figure 10. Guessed years of mixed and single-year cards. This figure compares the years guessed by ChatGPT-5.6 Sol with the source years of the information cards. Panel A plots the guessed year against the mean source year of the card's nine year-varying modules for each of the 288 mixed cards. The printed values are the standard deviation of the guesses and their correlation with that mean year. The fixed tenth module carries identical text on every card and has no source year. Panel B plots the guessed year against the true year of the 12 single-year benchmark cards, each built from one year's nine masked modules without mixing. The dashed line marks exact agreement.

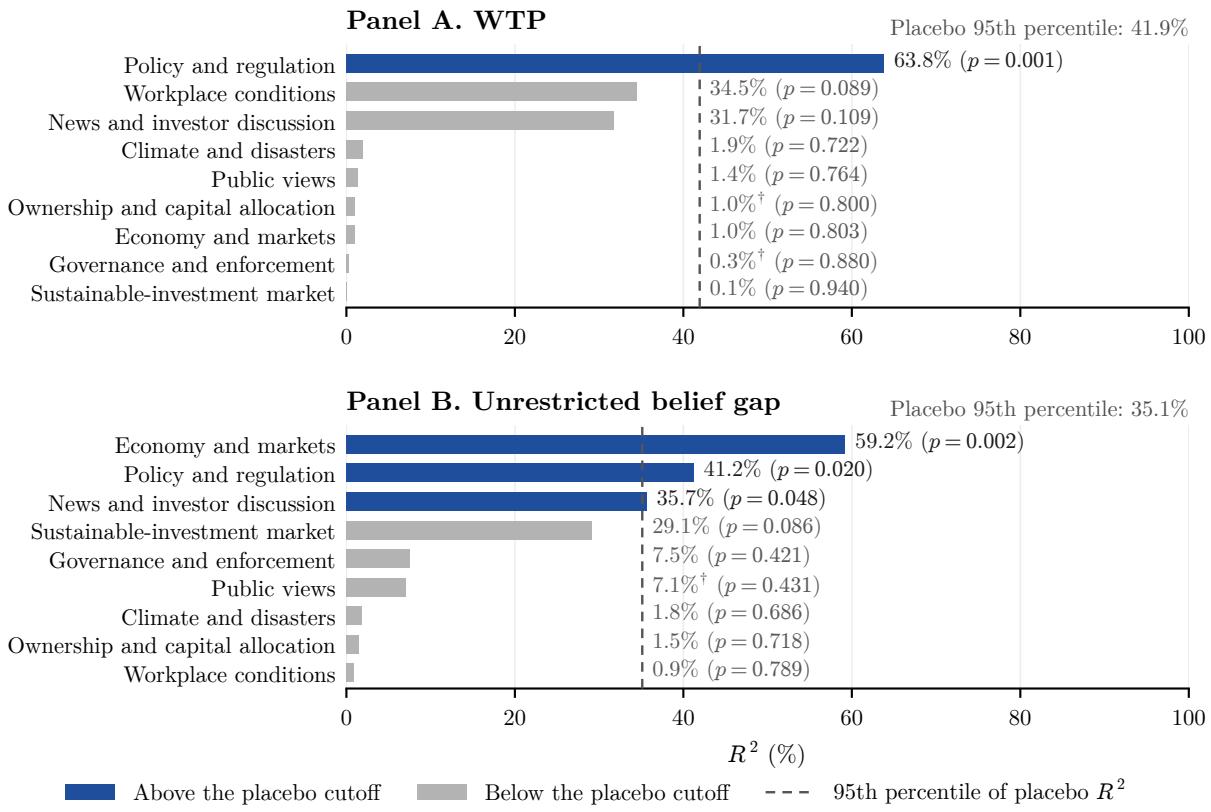


Figure 11. Module R^2 for reconstructed tastes and beliefs. This figure reports how closely each of the nine year-varying information modules tracks the aggregate reconstructed series over the 2014–2025 environments in ChatGPT-5.6 Sol. The measure is the R^2, in percent, of a univariate regression of the aggregate series on the module's annual effects. Each regression includes an intercept. Panel A reports willingness to pay (WTP), and Panel B reports the unrestricted belief gap. WTP averages the annual return given up for the otherwise matched high-score portfolio and the extra return required for the low-score portfolio over ten years. The gap is the expected annual return over ten years on typical real-world portfolios at the 80th ESG percentile minus that at the 20th. Module paths come from equation (3) with the fixed tenth module added as a tenth set of year indicators. The aggregate series is the sum of the ten paths and the intercept. It weights the seven identities by the fixed 2014 corporate-equity holdings of their sectors. Within each panel, modules are ordered by their R^2. Each bar is labeled with its R^2 and its placebo p-value. The p-value is the share of 1,000 random reassignments of the fixed module's year labels in which the relabeled module's R^2 is at least as large as the module's own. The dashed line marks the 95th percentile of the placebo R^2, 41.9% for WTP and 35.1% for the belief gap. Blue bars exceed it. A dagger marks a module whose effects are negatively correlated with the aggregate. Figure B.10 plots the placebo distributions.

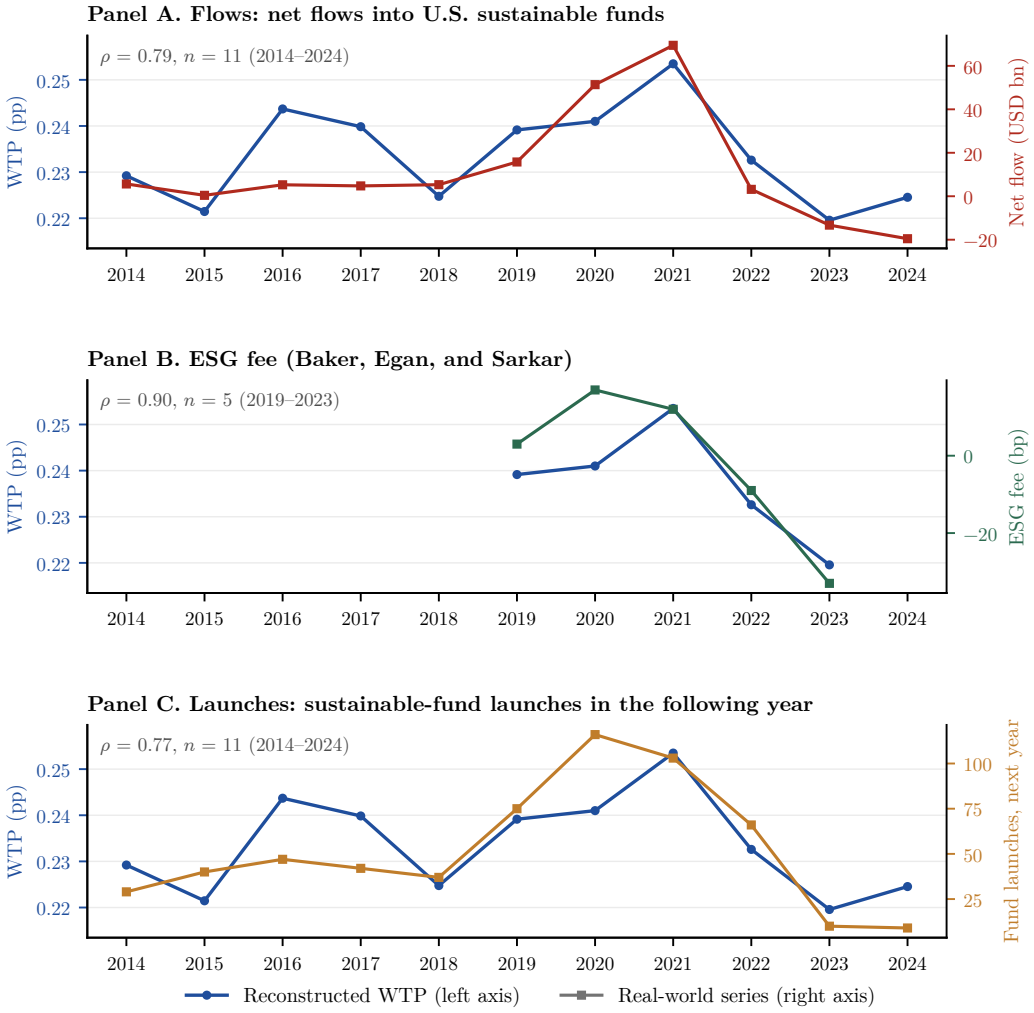


Figure 12. Real-world investing series and reconstructed tastes. This figure compares reconstructed willingness to pay (WTP) in ChatGPT-5.6 Sol with three annual series on sustainable investing. WTP averages the annual return given up for the high-score portfolio and the extra return required for the low-score portfolio over ten years; the otherwise identical portfolios score at the 80th and 20th ESG percentiles. Circles show the weighted WTP of Figure 9 on the left axis, in annual percentage points, over the years available for each real-world series. Squares show the real-world series on the right axis. Panel A plots net flows into U.S. sustainable funds, in USD billion (Morningstar). Panel B plots the ESG fee of Baker et al. (2024): the additional annual fee, in basis points, that investors are willing to pay for index funds with an ESG mandate. Panel C plots next-year sustainable-fund launches (Morningstar). Each panel prints the levels correlation ρ, the number of annual observations, and the sample period. The left axis is scaled to the range of WTP, 3.4 bps over 2014–2025. Table B.7 documents sources.

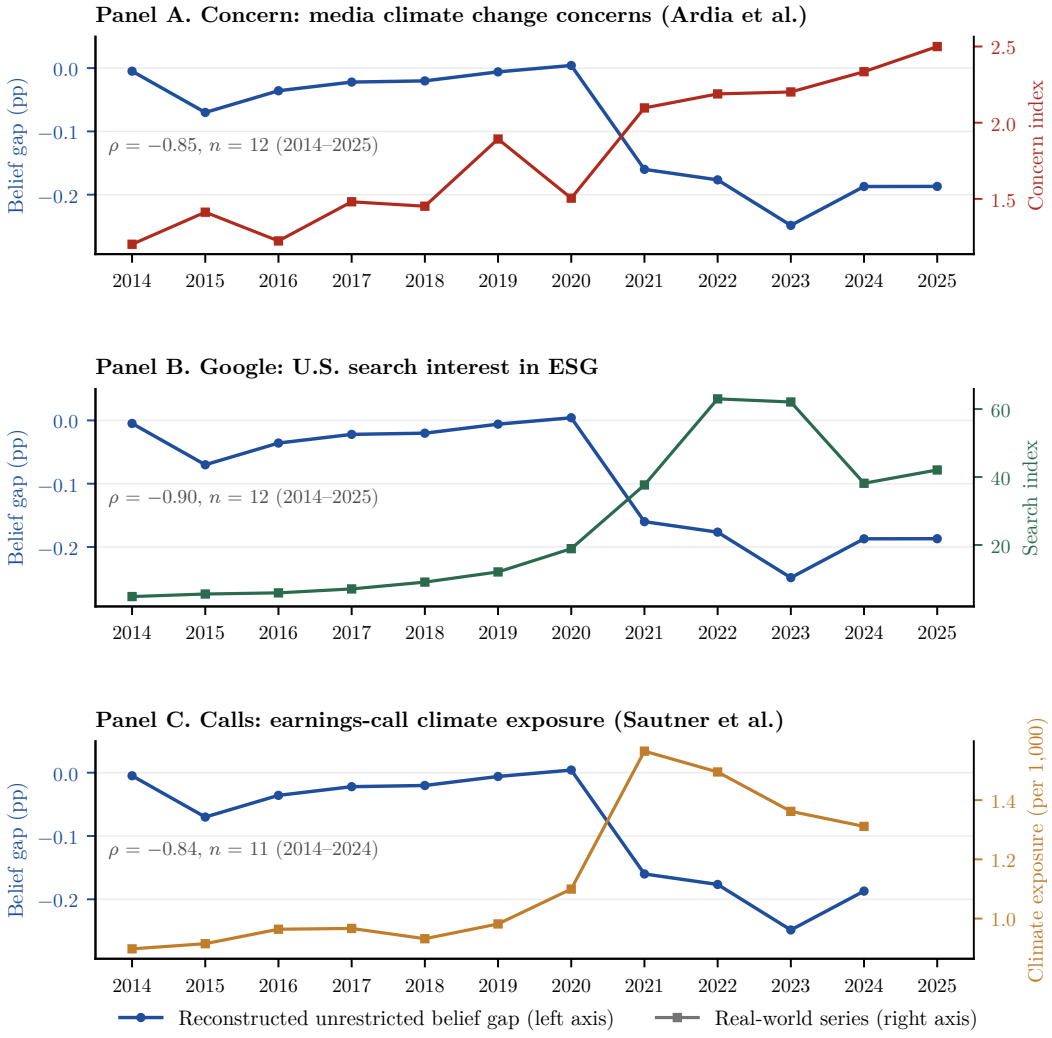


Figure 13. Real-world attention series and reconstructed beliefs. This figure compares the reconstructed unrestricted belief gap in ChatGPT-5.6 Sol with three annual measures of attention to climate and ESG. The gap is the expected annual return over ten years on typical real-world portfolios at the 80th ESG percentile minus that at the 20th. A falling gap means that green portfolios are expected to do worse relative to brown ones. Circles show the weighted gap of Figure 9 on the left axis, in annual percentage points. Squares show the real-world series on the right axis. Panel A plots the Media Climate Change Concerns index (Ardia et al., 2023), with 2025 equal to the January–June mean. Panel B plots U.S. Google search interest in ESG, in relative search volume. Panel C plots the climate-change exposure of U.S. firms' earnings calls (Sautner et al., 2023), in climate-change bigrams per thousand, over 2014–2024. Each panel prints the levels correlation ρ, the number of annual observations, and the sample period. The left axis is scaled to the range of the gap, 25.3 bps. Table B.7 documents sources.

Table 1. Summary of survey design

Design element	Count
Retail identities (4 age × 4 income × 5 politics × 2 gender × 3 literacy)	480
Institutional identities (5 type × 3 AUM × 4 geography)	60
Cells per identity (baseline, E, S, G, recession, boom)	6
Total cells	3,240
Question modules per cell (preference, matched beliefs, unrestricted beliefs, allocation, motivation)	5
ChatGPT-5.6 Sol (main model): all cells × 5 modules × 3 repeats	48,600
ChatGPT-5.5, Claude Opus 5.0, Claude Fable 5: all cells × 5 modules × 1 each	48,600
Nine further model versions: shared cell subsamples × 3 belief/preference modules each	7,200
Total calls	104,400

Notes: A cell consists of one identity under one condition, with each of the five question modules submitted in a separate model call. The nine further model versions answer the preference module and the two belief modules on fixed cell subsamples shared across models. All 104,400 calls returned usable answers.

Table 2. Retail attributes and WTP

	Coefficient	(SE)	t
Political affiliation (omitted: Neutral)			$\Delta R^2 = 0.751$
Strong Democrat	0.55^{***}	(0.028)	20.0
Weak Democrat	0.16^{***}	(0.015)	10.1
Weak Republican	-0.23^{***}	(0.013)	-18.1
Strong Republican	-0.47^{***}	(0.025)	-18.8
Income (omitted: 40th to 80th percentile)			$\Delta R^2 = 0.029$
Bottom 40%	-0.08^{***}	(0.018)	-4.6
80th to 95th percentile	0.05^{**}	(0.020)	2.3
Top 5%	0.10^{***}	(0.025)	4.2
Gender (omitted: male)			$\Delta R^2 = 0.023$
Female	0.12^{***}	(0.016)	7.8
Financial literacy (omitted: medium)			$\Delta R^2 = 0.014$
Low	0.08^{***}	(0.021)	3.8
High	-0.03^{**}	(0.015)	-2.2
Age (omitted: 45 to 64)			$\Delta R^2 = 0.009$
Under 35	0.09^{***}	(0.023)	3.9
35 to 44	0.01	(0.021)	0.3
65 or older	0.00	(0.021)	0.1
Constant (reference identity)	0.37^{***}	(0.024)	15.5
R^2			0.825
Identity cells			480

Notes: This table reports a regression of ChatGPT-5.6 Sol's willingness to pay (WTP) on category indicators for all five attributes at once: politics, income, gender, financial literacy, and age. The sample contains 480 baseline retail cells with three repeats averaged within each cell. WTP averages the annual return given up for Portfolio H and the extra return required for Portfolio L over ten years, in annual percentage points. The portfolios are otherwise identical and score at the 80th (H) and 20th (L) ESG percentiles; the economy is normal. Each coefficient is relative to the attribute's omitted category. The constant represents a politically Neutral male aged 45–64, with income in the 40th–80th percentile and medium financial literacy. Standard errors in parentheses are heteroskedasticity-robust and summarize variation in the model's answers across the designed identity grid. ***, **, and * denote $|t| \geq 2.58$, 1.96, and 1.645, respectively. ΔR^2 is the decline in R^2 when the attribute's indicators are removed from the full model. These variance shares describe the designed synthetic grid and should not be interpreted as estimates for a human population; Section 7.1 discusses response compression.

Table 3. Institutional attributes and WTP

	Coefficient	(SE)	t
Institution type (omitted: pension fund)			$\Delta R^2 = 0.616$
Endowment or foundation	0.09^{***}	(0.014)	6.7
Insurance company	-0.06^{***}	(0.012)	-4.7
Hedge fund	-0.08^{***}	(0.014)	-5.9
Mutual fund company	-0.09^{***}	(0.012)	-7.7
Geography (omitted: U.S.)			$\Delta R^2 = 0.253$
Europe	0.12^{***}	(0.011)	11.1
Asia	0.06^{***}	(0.012)	4.8
Other regions	0.07^{***}	(0.012)	5.7
AUM size (omitted: medium)			$\Delta R^2 = 0.023$
Small	-0.03^{***}	(0.010)	-3.1
Large	-0.02^{**}	(0.010)	-2.3
Constant (reference institution)	0.16^{***}	(0.012)	13.3
R^2			0.892
Identity cells			60

Notes: This table reports a regression of ChatGPT-5.6 Sol's willingness to pay (WTP) on category indicators for all three attributes at once: institution type, geography, and assets under management (AUM). The sample contains 60 baseline institutional cells with three repeats averaged within each cell. WTP averages the annual return given up for Portfolio H and the extra return required for Portfolio L over ten years, in annual percentage points. The portfolios are otherwise identical and score at the 80th (H) and 20th (L) ESG percentiles; the economy is normal. Each coefficient is relative to the attribute's omitted category, and the constant represents a medium-size U.S. pension fund. Standard errors in parentheses are heteroskedasticity-robust and summarize variation in the model's answers across the designed identity grid. ***, **, and * denote $|t| \geq 2.58$, 1.96, and 1.645, respectively. ΔR^2 is the decline in R^2 when the attribute's indicators are removed from the full model.

Table 4. The stated portfolio at baseline

	ChatGPT-5.6 Sol	ChatGPT-5.5	Claude Opus 5.0	Claude Fable 5
Retail (480 identities)				
Share of wealth in U.S. stocks	52.1 (0.64)	52.3 (0.66)	54.8 (0.60)	51.6 (0.72)
Share of stocks in Portfolio H	60.2 (0.98)	60.4 (1.03)	54.0 (0.61)	57.4 (0.96)
Implied share of wealth in Portfolio H	31.7 (0.68)	31.7 (0.70)	29.6 (0.47)	29.9 (0.70)
Institutions (60 identities)				
Share of AUM in U.S. public equities	23.4 (1.51)	20.1 (1.55)	27.7 (2.09)	25.5 (1.95)
Share of equities in Portfolio H	70.4 (1.22)	79.7 (1.44)	63.3 (1.34)	68.1 (1.47)
Implied share of AUM in Portfolio H	15.8 (0.85)	15.4 (1.00)	16.7 (1.12)	16.5 (1.12)

Notes: The table reports mean stated allocations across 480 retail and 60 institutional baseline identity cells, in percent. The baseline uses aggregate ESG scores and a normal economy. Repeats are averaged within each cell; standard errors across cells are in parentheses and summarize variation in the model's answers across the designed identity grid. Portfolios H and L are otherwise identical and score at the 80th and 20th percentiles of the same greenness measure. The hypothetical allocation question offers H and L as the only equity choices. For retail identities, the two shares are financial wealth in U.S. stocks and the stock allocation in H; for institutions, they are assets under management (AUM) in U.S. public equities and the equity allocation in H. The implied share of wealth or AUM in H multiplies the two stated shares within each answer before averaging.

Table 5. Portfolio H share, WTP, and the matched belief gap

	ChatGPT-5.6 Sol			ChatGPT-5.5			Claude Opus 5.0			Claude Fable 5		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
WTP	39.15***		26.29***	44.87***		30.75***	21.41***		17.26***	40.00***		34.31***
	(0.87)		(1.06)	(1.00)		(1.08)	(0.16)		(0.39)	(0.36)		(0.75)
Matched belief gap		26.70***	13.72***		23.30***	12.49***		12.12***	2.89***		13.30***	2.34***
		(0.45)	(0.61)		(0.44)	(0.47)		(0.13)	(0.24)		(0.18)	(0.28)
Fixed effects	No	No	No	No	No	No	No	No	No	No	No	No
R^2	0.50	0.44	0.57	0.48	0.42	0.56	0.79	0.68	0.80	0.78	0.66	0.79
Cells	3,240	3,240	3,240	3,240	3,240	3,240	3,240	3,240	3,240	3,240	3,240	3,240

Notes: The dependent variable is the stated share of the equity allocation assigned to Portfolio H, in percent. Each column reports one OLS regression with an intercept, which is not shown. For each model, the first column uses willingness to pay (WTP) alone, the second the matched belief gap alone, and the third both. WTP averages the return given up for H and the extra return required for L, and the matched belief gap is the expected return on H minus that on L. Both are in annual percentage points over ten years, so each coefficient is the change in the Portfolio H share, in percentage points, per percentage point of the regressor. H and L are otherwise identical portfolios at the 80th and 20th percentiles of the same greenness measure and are the allocation question's only equity choices. Each regression uses all 3,240 cells of the model: 480 retail and 60 institutional identities under six score and economy conditions. For ChatGPT-5.6 Sol, the three repeats are averaged within each cell. WTP, beliefs, and allocations are elicited in separate model calls. Standard errors in parentheses are heteroskedasticity-robust and summarize variation in the model's answers across the designed identity grid. ***, **, and * denote $|t| \geq 2.58$, 1.96, and 1.645, respectively.

Table 6. Validation: marginal distribution matching

	Giglio et al. (2025)	ChatGPT-5.6 Sol	Claude Fable 5
Expected market return, mean (SD)	7.13 (3.96)	6.75 (0.33)	6.12 (0.51)
Expected ESG return, mean (SD)	5.20 (4.88)	6.19 (0.44)	5.21 (0.68)
ESG-minus-market belief gap, mean	−2.06	−0.56	−0.91
Belief-gap SD	5.34	0.31	0.51
Share expecting ESG underperformance (%)	–	99.2	99.3
Democratic-county belief gap	−1.86	−0.38	−0.58
Republican-county belief gap	−2.40	−0.93	−1.58
Female belief gap	−1.75	−0.48	−0.77
Male belief gap	−2.19	−0.61	−0.99
High climate-concern belief gap	−1.04	−0.38	−0.58
Moderate climate-concern belief gap	−2.17	−0.85	−1.33
Low climate-concern belief gap	−4.82	−1.04	−1.60
Belief gap by motive for ESG investing:			
"ESG will outperform"	1.00	–	–
"ESG hedges climate risk"	−0.65	−0.41	−0.43
"It is the right thing to do"	−0.98	−0.41	−0.58
"No specific reason"	−3.64	−0.99	−1.58
Responses	varies by row	1,000 profiles	1,000 profiles

Notes: The table compares the human survey benchmark with responses from ChatGPT-5.6 Sol and Claude Fable 5 to the same questions for the same 1,000 synthetic profiles. Profiles match the published sample's marginal distributions of age, wealth, gender, county political majority, and flood risk, one characteristic at a time. The ESG questions reproduce Figure 1 of Giglio et al. (2025) verbatim; the aggregate market-return question reproduces the survey appendix of Giglio et al. (2021) verbatim. Both models receive an instruction to answer as of December 2023, the final published wave, without a restriction on training information. The human benchmark pools waves from June 2021 through December 2023; climate-concern questions cover thirteen waves beginning in December 2021. Expected returns and ESG-minus-market belief gaps are annual percentage points over ten years. Response samples differ across rows; the separately reported market and ESG means therefore need not yield the paired belief gap. The published table reports a median gap of −1 but no share expecting underperformance. Motive rows report mean belief gaps conditional on the selected motive. Neither model selects "ESG will outperform," leaving those two cells empty.

Table 7. Validation: joint distribution matching

	ChatGPT-5.6 Sol		Claude Fable 5	
	mean	(SE)	mean	(SE)
Expected market return	6.17	(0.018)	5.92	(0.027)
Expected ESG return	5.57	(0.019)	4.81	(0.024)
ESG-minus-market belief gap	−0.60	(0.004)	−1.11	(0.010)
Share of profiles with a negative gap (%)	98.5		99.8	

Notes: The table reports synthetic responses for all 4,595 households in the public-use file of the 2022 Survey of Consumer Finances (SCF). ChatGPT-5.6 Sol and Claude Fable 5 each answer the four questions in Table 6 once per profile: the three ESG questions of Giglio et al. (2025) and the market-return question of Giglio et al. (2021). Profiles preserve each household's joint characteristics, whereas Table 6 matches marginal distributions separately. The instruction to answer as of 2022 imposes no restriction on training information. Means and standard errors use SCF sampling weights. Expected returns and ESG-minus-market belief gaps are annual percentage points over ten years; the share with a negative gap is in percent.

Table 8. Real-world series and the reconstructed tastes and beliefs

	Investing			Attention		
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Flows	Fee	Launches	Concern	Google	Calls
WTP	0.78***	0.70*	0.74**	0.21	0.06	0.46***
	(3.46)	(4.10)	(3.15)	(1.27)	(0.44)	(3.74)
Beliefs	0.02	0.44	0.11	−0.88***	−0.91***	−0.96***
	(0.11)	(2.57)	(0.45)	(−5.37)	(−6.37)	(−7.91)
Observations	11	5	11	12	12	11
R^2	0.62	0.95	0.60	0.76	0.82	0.89
Sample	2014–24	2019–23	2014–24	2014–25	2014–25	2014–24

Notes: Each column reports a regression of an annual real-world series on the reconstructed willingness to pay (WTP) and the reconstructed unrestricted belief gap (Beliefs), the weighted ChatGPT-5.6 Sol series of Figure 9. Columns (1)–(3) are investing series and columns (4)–(6) attention series. WTP averages the annual return given up for an otherwise matched high-ESG portfolio and the extra return required for its low-ESG counterpart; Beliefs is the expected return on typical real-world high-ESG portfolios minus that on low-ESG portfolios. Both use the 80th and 20th score percentiles and annual returns over ten years. All three variables are standardized to mean zero and standard deviation one within the column's sample, so coefficients are in standard-deviation units. Parentheses report t-statistics using OLS standard errors. ***, **, and * denote two-sided p-values below 0.01, 0.05, and 0.10 from the t distribution with $n - 3$ degrees of freedom. Flows are net flows into U.S. sustainable funds, in USD billion (Morningstar). Fee is the additional annual fee, in basis points, that investors are willing to pay for index funds with an ESG mandate, estimated from index-fund flows by Baker et al. (2024). Launches is the number of sustainable funds launched in the following year (Morningstar). Concern is the Media Climate Change Concerns index of Ardia et al. (2023), built from climate-change articles in major U.S. newspapers and newswires. Google is U.S. Google search interest in ESG, in relative search volume. Calls is the mean climate-change exposure of U.S. firms' earnings calls, in climate-change bigrams per thousand (Sautner et al., 2023). Samples are 2014–2024 for Flows, Launches, and Calls, 2019–2023 for Fee, and 2014–2025 for Concern and Google. The 2019 Fee estimate covers May–December, and the 2025 Concern observation is the January–June mean. Flows use the January 2025 Morningstar vintage; launches use that vintage through 2024 and the March 2026 review for the 2025 count. Table B.7 gives source definitions, and Tables B.8 and B.9 add sentiment controls.

Internet Appendix

A Survey Instruments

This appendix documents the survey instruments, including the identity descriptions, question modules, value and economy variants, validation questionnaire, and dated-sample instruments. We refer to the preference and belief calls as the main round, the allocation and motivation calls as the allocation round, and the calls in Sections 7.1 and 8 as the validation and dated rounds. Each call to an investor respondent has two parts: a system message with the respondent identity and the standing instructions, and a user message with the questions. The year-guessing probe uses a separate analyst instruction, reproduced in Appendix A.4. All calls were run as fresh conversations, with no prior conversation history, worked examples, or tools. The message blocks below, together with the identity records and information cards in the replication package, show how the calls were constructed.

The appendix reproduces message text together with the templates and assembly rules used to construct the calls. Bracketed fields denote substitutions made by the code; their definitions and permitted values are provided with the corresponding templates. Bracketed category labels, SCF age and income codes, formatting metadata, and assembly notation are explanatory annotations rather than literal parts of the investor messages. Repeated question blocks are reproduced once and referenced where they recur. Wording, punctuation, and source line breaks are preserved within the reproduced message blocks. Long lines wrap with a hanging indent so that typesetting wraps can be distinguished from line breaks in the source text.

A.1 Identity Descriptions

A respondent identity is a set of declarative sentences that opens the system message. The design grid crosses the attributes below to produce the 480 retail and 60 institutional identities.

Retail respondents

The retail template carries five attributes: age, household income, political behavior, gender, and financial literacy.

```
You are an individual retail investor in the United States, taking part in an investor
    survey. Your age group is [AGE GROUP]. Your annual household income is in [INCOME
    GROUP]. [POLITICAL-BEHAVIOR SENTENCE] Your gender is [GENDER]. Your financial
    literacy is [FINANCIAL LITERACY], meaning [FINANCIAL-LITERACY DEFINITION].
```

The age and income categories group the SCF's bins (Board of Governors, 2023), as shown below.

Age group.

```
under 35   (SCF codes: AGECL 1; definition: you are younger than 35)
35 to 44   (SCF codes: AGECL 2; definition: you are between 35 and 44)
45 to 64   (SCF codes: AGECL 3-4; definition: you are between 45 and 64)
65 or older   (SCF codes: AGECL 5-6; definition: you are 65 or older)
```

Household income group.

```
the bottom 40 percent of the U.S. household income distribution
    (SCF codes: INCPCTLECAT 1-4; definition: below the 40th percentile of U.S.
    household income)
the 40th to 80th percentile of the U.S. household income distribution
    (SCF codes: INCPCTLECAT 5-8; definition: between the 40th and 80th percentiles of
    U.S. household income)
the 80th to 95th percentile of the U.S. household income distribution
    (SCF codes: INCPCTLECAT 9-10; definition: between the 80th and 95th percentiles of
    U.S. household income)
the top 5 percent of the U.S. household income distribution
    (SCF codes: INCPCTLECAT 11-12; definition: above the 95th percentile of U.S.
    household income)
```

Political behavior. Party registration and donation behavior are stated explicitly. The bracketed labels name the five levels in our design and are not shown to the model.

```
[Strong Democrat]
You are a registered Democratic voter and have donated to the Democratic Party in each
    of the last three years.

[Weak Democrat]
You are a registered Democratic voter but have not donated to the Democratic Party in
    the last three years.
```

```
[Neutral]
You are not registered with either major party, have not donated to either party, and
    do not lean consistently toward either party.

[Weak Republican]
You are a registered Republican voter but have not donated to the Republican Party in
    the last three years.

[Strong Republican]
You are a registered Republican voter and have donated to the Republican Party in each
    of the last three years.
```

Gender.

```
female
male
```

Financial literacy. The category name fills [FINANCIAL LITERACY], and the corresponding description fills [FINANCIAL-LITERACY DEFINITION].

```
[low]
you would answer zero or one of the standard three financial-literacy questions
    correctly

[medium]
you would answer two of the standard three financial-literacy questions correctly

[high]
you would answer all three of the standard three financial-literacy questions correctly
```

Institutional respondents

The institutional template specifies institution type, assets under management, geography, and the mandate associated with the institution type. The code supplies the appropriate article before the institution type, as in "an endowment or foundation." The opening sentence refers to an institutional investor "in the United States" for all profiles, including those whose later geography clause specifies non-U.S. headquarters. Both clauses are reproduced as administered.

```
You are the investment decision-maker at an institutional investor in the United
    States, taking part in an investor survey. Your institution is a/an [INSTITUTION
    TYPE]. Your mandate is to [MANDATE]. Your assets-under-management category is [AUM
    SIZE], that is, [AUM RANGE]. Your institution is [GEOGRAPHY DEFINITION].
```

Institution type and mandate.

```
[insurance company]
manage the general account against insurance liabilities, regulatory-capital
    constraints, liquidity needs, and policyholder obligations

[endowment or foundation]
support a spending rule while preserving the real long-run purchasing power of the fund

[pension fund]
meet long-dated beneficiary liabilities under fiduciary, funding-ratio, and liquidity
    constraints

[hedge fund]
maximize risk-adjusted performance subject to the fund strategy, investor terms,
    leverage, and redemptions

[mutual fund company]
follow the stated prospectus and benchmark while meeting fiduciary duties and
    daily-liquidity obligations
```

Assets under management.

```
[small]
less than USD 5 billion in total assets under management

[medium]
USD 5 billion to less than USD 50 billion in total assets under management

[large]
at least USD 50 billion in total assets under management
```

Geography. The bracketed labels identify the categories below. The prompt uses the text under each label to describe the institution's headquarters and its beneficiary or client base.

```
[US]
headquartered in the United States, with a majority U.S. beneficiary or client base

[Europe]
headquartered in Europe, with a majority European beneficiary or client base

[Asia]
headquartered in Asia, with a majority Asian beneficiary or client base

[other regions]
headquartered outside the United States, Europe, and Asia, with the majority
    beneficiary or client base in that same region
```

Respondent stance

The two stance paragraphs below, one for each investor class, ask for the respondent's own view rather than a category average or an expert consensus. Each is used unchanged in every module.

Retail.

```
Answer as this person, not as an analyst, a financial adviser, a representative for
    any group, or an average investor. Your answers should reflect your own views,
    including any disagreement with expert or consensus opinion. Investors who share
    these characteristics do not all answer alike, so answer as this particular
    respondent rather than reporting a category average. Give a point estimate for
    each question, and do not mechanically report round or focal numbers.
```

Institutional.

```
Answer for this specific institution, not as an outside analyst, a consultant, an
    industry association, or a representative institution of its category. Your
    answers should reflect this institution's own mandate, beneficiaries or clients,
    governance, and investment culture, including any disagreement with expert or
    consensus opinion. Institutions in the same category do not all answer alike, so
    answer for this particular institution rather than reporting a category average.
```

```
Give a point estimate for each question, and do not mechanically report round or
focal numbers.
```

A.2 Question Modules

For each identity-condition cell, the five modules below are administered in separate calls, each containing two questions. A call includes no questions or answers from another module. Responses are then linked by identity and condition.

The two portfolios

All five modules use the same score sentence, which places Portfolios H and L at the 80th and 20th percentiles of one scoring methodology. The preference, matched-belief, allocation, and motivation modules also use the matching paragraph below. The unrestricted-belief module omits it.

Score sentence.

```
Under one fixed [VALUE LABEL] scoring methodology and date, Portfolio H has a
    weighted-average [VALUE LABEL] score at the 80th percentile and Portfolio L has a
    weighted-average [VALUE LABEL] score at the 20th percentile of the same
    cross-section of publicly traded U.S. stocks.
```

Matching paragraph.

```
The stated [VALUE LABEL] score and the practices directly captured by that score are
    the only differences between the two portfolios. Every holding in Portfolio H has
    a same-weight matched counterpart in Portfolio L. Hold all other financial and
    non-financial characteristics constant now and throughout the next 10 years,
    including industry mix, cash-flow prospects, growth, risk, size, capital
    structure, valuation, factor exposures, diversification, concentration, turnover,
    fees, liquidity, construction rule, and rebalancing rule.
```

Module 1. Matched preference

This is the willingness-to-pay question, asked from both sides: the return given up to hold Portfolio H, and the extra return required to hold Portfolio L. The closing clause of the setup allows answers of either sign.

Setup.

```
Here, [VALUE LABEL] covers [VALUE TOPIC].

Consider two hypothetical, long-only, broadly diversified portfolios of publicly
    traded U.S. stocks, Portfolio H and Portfolio L. [SCORE SENTENCE]

[MATCHED-CONDITION PARAGRAPH]

Treat the two questions below as hypothetical return tradeoffs over the next 10 years.
    Base each answer only on [PREFERENCE BASIS]. Do not forecast returns. Your answers
    may be positive, zero, or negative.
```

The basis clause, retail.

```
your own non-pecuniary values and the stated [VALUE LABEL] difference
```

The basis clause, institutional.

```
your institution's own non-pecuniary values, its mandate, and its beneficiaries' or
    clients' preferences, together with the stated [VALUE LABEL] difference
```

High-side willingness-to-pay question.

```
How many percentage points of average annual return would you be willing to give up in
    order to hold Portfolio H rather than Portfolio L over the next 10 years, based
    only on the stated [VALUE LABEL] difference? A positive number means you would
    accept a lower return from Portfolio H; zero means you would be indifferent if the
    two returns were equal; a negative number means Portfolio H would have to offer
    you a higher return.
```

Low-side required-compensation question.

```
How many percentage points of additional average annual return would you require in
    order to hold Portfolio L rather than Portfolio H over the next 10 years, based
    only on the stated [VALUE LABEL] difference? A positive number means Portfolio L
```

```
would have to offer you a higher return; zero means you would be indifferent if
    the two returns were equal; a negative number means you would accept a lower
    return from Portfolio L.
```

The order of the two questions is determined by a fixed hash-based rule. For each cell and module, we compute the SHA-256 hash of the string `20260802|cell_id|module`, convert its first eight hexadecimal digits to an integer, and present H first if the integer is even and L first otherwise. The order remains fixed across models and repeated calls for the same cell and module.

Module 2. Matched belief

The matched-belief module retains the identity, portfolio descriptions, and matching paragraph. It replaces the preference setup and questions with expected-return questions and uses its own system instruction and answer keys.

Setup, replacing the preference module's closing paragraph.

```
Given that matching, report what you expect each portfolio to earn. Report the
    expected average annual total return, including dividends and capital gains or
    losses, in percent per year over the next 10 years. Your answers may be positive,
    zero, or negative.
```

The two questions. In the user message, the labels `Portfolio H forecast:` and `Portfolio L forecast:` precede the corresponding questions, each on a separate line.

```
What average annual total return do you expect Portfolio H to earn over the next 10
    years?
What average annual total return do you expect Portfolio L to earn over the next 10
    years?
```

Module 3. Unrestricted belief

The unrestricted-belief module omits the matching paragraph and asks about typical real-world portfolios. It instructs the model to consider the characteristics that usually accompany each greenness score.

Complete setup.

```
Here, [VALUE LABEL] covers [VALUE TOPIC].

This is a real-world forecasting question about two typical, long-only, broadly
    diversified portfolios of publicly traded U.S. stocks, Portfolio H and Portfolio
    L. [SCORE SENTENCE]

Report the expected average annual total return, including dividends and capital gains
    or losses, in percent per year over the next 10 years. Your answers may be
    positive, zero, or negative. Use the real-world features that typically accompany
    the score level of each corresponding portfolio.
```

The two questions and their labels are identical to those of Module 2.

Module 4. Allocation

The allocation module uses the same identities and design grid as the main round and retains the matched-portfolio setup. Its own instruction, closing setup paragraph, questions, and answer keys elicit two allocation margins in one call: the share assigned to U.S. equities and the share of that equity allocation assigned to Portfolio H.

Closing setup paragraph.

```
The two questions below ask about the portfolio you would actually hold over the next
    10 years. Base each answer only on [BASIS]. Both answers are percentages between 0
    and 100, but they are percentages of different things: the first is a share of
    everything you hold, and the second is a share of the U.S. stock part alone. Each
    question states its own base.
```

The basis clause, retail.

```
your own financial circumstances and your own non-pecuniary values
```

The basis clause, institutional.

```
your institution's own non-pecuniary values, its mandate, its liabilities, its
    governance process, and its beneficiaries' or clients' preferences
```

Participation margin, retail.

```
Out of your total financial wealth, what percentage would you invest in the U.S. stock
    market? The remainder is held in everything else, such as cash, bonds, and real
    estate. Answer as a percentage of your total financial wealth, between 0 and 100.
```

Composition margin, retail.

```
Now take that stock-market wealth as given, and suppose it is invested entirely in
    these two portfolios, so that whatever does not go to Portfolio H goes to
    Portfolio L, and the share in Portfolio H and the share in Portfolio L add up to
    100. Out of that stock-market wealth, what percentage would you hold in Portfolio
    H? Answer as a percentage of your stock-market wealth, between 0 and 100, where 0
    means all of it is in Portfolio L, 50 means an even split, and 100 means all of it
    is in Portfolio H.
```

Participation margin, institutional.

```
Out of your institution's total assets under management, what percentage would it hold
    in publicly traded U.S. stocks? The remainder is held in everything else it
    invests in, such as bonds, private assets, and cash. Answer as a percentage of
    total assets under management, between 0 and 100.
```

Composition margin, institutional.

```
Now take that U.S. public-equity allocation as given, and suppose it is invested
    entirely in these two portfolios, so that whatever does not go to Portfolio H goes
    to Portfolio L, and the share in Portfolio H and the share in Portfolio L add up
    to 100. Out of that U.S. public-equity allocation, what percentage would your
    institution hold in Portfolio H? Answer as a percentage of the U.S. public-equity
    allocation, between 0 and 100, where 0 means all of it is in Portfolio L, 50 means
    an even split, and 100 means all of it is in Portfolio H.
```

The questions are labeled "Question 1 (field name):" and "Question 2 (field name):", with the field name identifying the JSON key for the answer.

Module 5. Motivation

In a separate call, the respondent divides a hypothetical holding of Portfolio H between own choice and outside pressure. The call does not include the respondent's allocation answer. The pressure descriptions differ by investor type: friends, family, colleagues, and other people

for retail identities; mandates, the investor base, trustees or clients, regulatory or industry expectations, and reputational considerations for institutions.

Holding sentence, retail.

```
Suppose you hold some of Portfolio H in your stock-market wealth.
```

Holding sentence, institutional.

```
Suppose your institution holds some of Portfolio H in its U.S. public-equity
    allocation.
```

Closing setup sentence, appended to the holding sentence.

```
The two questions below divide that holding into two parts. The two percentages must
    add up to 100.
```

Own-choice question, retail.

```
Of that Portfolio H holding, what percentage is there because you personally wanted
    it, for your own reasons? Answer as a percentage between 0 and 100.
```

Outside-pressure question, retail.

```
Of that same Portfolio H holding, what percentage is there because of pressure from
    other people, such as what your friends, family, colleagues, or people whose
    opinion you care about would think, rather than your own reasons? Answer as a
    percentage between 0 and 100.
```

Own-choice question, institutional.

```
Of that Portfolio H holding, what percentage is there because your institution
    actively chose it within its own discretion, as an allocation it wanted to make on
    its own judgment? Answer as a percentage between 0 and 100.
```

Outside-pressure question, institutional.

```
Of that same Portfolio H holding, what percentage is there because of your
    institution's mandate, its investor base, its trustees or clients, regulatory or
    industry expectations, or reputational considerations, rather than its own
    judgment? Answer as a percentage between 0 and 100.
```

Value dimensions

The aggregate and single-pillar variants differ only in the value label and topic list, drawn from the MSCI theme taxonomy (MSCI, 2026). MSCI ratings assess industry-relative financial risks and opportunities, whereas the score of the matched portfolios in our survey is hypothetical. All other instrument wording is identical across aggregate ESG and the environmental, social, and governance variants, apart from the value-label substitutions elsewhere in the message.

```
[aggregate_esg]
label = aggregate ESG
topic = all ten MSCI March 2026 ESG themes: Climate Change, Natural Capital, Pollution
    & Waste, Environmental Opportunities, Human Capital, Product Liability,
    Stakeholder Opposition, Social Opportunities, Corporate Governance, Corporate
    Behavior

[environmental]
label = Environmental
topic = the MSCI March 2026 Environmental pillar themes: Climate Change, Natural
    Capital, Pollution & Waste, Environmental Opportunities

[social]
label = Social
topic = the MSCI March 2026 Social pillar themes: Human Capital, Product Liability,
    Stakeholder Opposition, Social Opportunities

[governance]
label = Governance
topic = the MSCI March 2026 Governance pillar themes: Corporate Governance, Corporate
    Behavior
```

The three single-pillar variants as delivered.

```
Here, Environmental covers the MSCI March 2026 Environmental pillar themes: Climate
    Change, Natural Capital, Pollution & Waste, Environmental Opportunities.

Here, Social covers the MSCI March 2026 Social pillar themes: Human Capital, Product
    Liability, Stakeholder Opposition, Social Opportunities.
```

```
Here, Governance covers the MSCI March 2026 Governance pillar themes: Corporate
    Governance, Corporate Behavior.
```

Economy states

Each system message in the main and allocation rounds includes one of the three economy-state descriptions below. Each description specifies the initial state and allows conditions to change over the ten-year horizon.

```
[normal]
At the time you answer, U.S. economic conditions are normal. Do not assume that this
    starting condition must persist for all ten years.

[recession_weak]
At the time you answer, the United States is in a recession, with weak growth,
    elevated macroeconomic risk, and cautious sentiment. Conditions may evolve or
    mean-revert over the ten-year horizon.

[boom_strong]
At the time you answer, the United States is in a boom, with strong growth, favorable
    sentiment, and relatively high risk appetite. Conditions may evolve or mean-revert
    over the ten-year horizon.
```

Standing instructions and system-message assembly

In addition to the identity and economy-state descriptions, each system message contains a module-specific instruction, an information-vintage instruction, and an output instruction. The module-specific instruction states whether the requested response is a return tradeoff, an expected return, an allocation, or a motivation share.

Module rules, main round.

```
[preference]
These two questions ask what you would be willing to give up, not what you forecast.
    Report your own return tradeoff. Do not report what you expect to happen, or give
    an answer you think is expected of you.

[belief_matched]
```

```
These two questions ask what you expect to happen, not what you prefer. Report your
    own expectations. Do not report what you would like to be true, or give an answer
    you think is expected of you.

[belief_unrestricted]
These two questions ask what you expect to happen, not what you prefer. Report your
    own expectations. Do not report what you would like to be true, or give an answer
    you think is expected of you.
```

Module rules, allocation round.

```
[allocation]
These two questions ask what you would actually hold, not what you forecast and not
    what you prefer. Report your own portfolio choice. Do not report what you think a
    well-diversified investor should hold, or give an answer you think is expected of
    you.

[motivation]
These two questions ask you to divide one holding you have already chosen into the
    part you want for your own reasons and the part you hold because of outside
    pressure. Report your own honest split. Do not report the split you think reflects
    well on you, or give an answer you think is expected of you.
```

Vintage rule and output rule, main and allocation rounds.

```
Answer using what you already know at the time you are asked.

Do not browse, cite sources, or explain your answers. Return only the requested JSON
    object.
```

Assembly. The blocks are joined by single spaces in this fixed order.

```
[identity sentences, 6 for retail / 5 for institutions]
    + ECONOMY_DESCRIPTIONS[economy_state]
    + RETAIL_PERSPECTIVE or INSTITUTION_PERSPECTIVE
    + MODULE_RULES[module]
    + VINTAGE_RULE
    + OUTPUT_RULE
```

Answer format

Each module ends its user message with an answer-format block specifying two JSON keys and requesting numeric answers with at most one decimal place. When a model's archived request configuration enables structured output, the call uses a matching schema. Otherwise the prompt alone specifies the format. The schema restricts key names and value types, while the decimal-place requirement is stated in the prompt. The preference block is shown below. The other modules use `belief_matched_high` and `belief_matched_low`; `belief_unrestricted_high` and `belief_unrestricted_low`; `equity_share_of_wealth` and `high_score_share_of_equity`; and `own_choice_share` and `outside_pressure_share`, respectively.

```
Return only valid JSON with exactly these keys, using JSON numbers with at most one
    decimal place.
{
  "preference_high": 0.0,
  "preference_low": 0.0
}
Do not return strings, explanations, confidence scores, markdown, additional keys, or
    any text before or after the JSON object.
```

A.3 Validation Instruments

The validation round uses the published survey questions unchanged. The three ESG questions come from Figure 1 of Giglio et al. (2025), and the market-return question from Online Appendix B.3 of Giglio et al. (2021). The code checks the questionnaire text against a stored fingerprint and rejects a run if the text has changed. The identity descriptions, respondent stance, user-message wrapper, and delivery formats are specific to our study. The formats replace the original web form's answer boxes and radio buttons.

The published questions

Aggregate-market expected return.

```
What do you expect the average annual return of the US stock market to be over the
    next 10 years?
```

Note: This expected return is the change in value, in percent, that you expect to receive each year on average over the next 10 years from investing in a portfolio that holds all stocks listed on the US stock market. It includes both dividends and capital gains/losses.

Answer note and unit label as printed beside the input box.

(Please answer only with a positive or negative numeric value, with at most 1 decimal.)

% per year, over the next 10 years

Definition of ESG investing, the preamble to the three ESG questions.

ESG investing offers a way to invest in funds with a core focus on environmental, social and governance issues. You may hear the term used interchangeably with "socially responsible investing" and "sustainable investing." Environmental issues involve protecting the natural environment; social issues relate to the firms' relationships with employees, suppliers, clients and other stake holders; and governance issues consider standards for company leadership (e.g. pay, board independence), risk controls and shareholder rights.

For these questions we would like to know your views on ESG equity investing, that is, investing in equity portfolios that score high on one or more of these dimensions.

Expected ESG return and answer instruction. The numeric-answer instructions for the market-return and ESG-return questions differ in punctuation. Each is reproduced with its original wording and punctuation.

What do you expect the average annual return of a diversified U.S. ESG equity portfolio to be over the next 10 years?

(Please answer only with a positive or negative numeric value with at most 1 decimal.)

Most important reason for ESG investing, and its four options in this order.

Listed below are some reasons why individuals might invest in ESG portfolios. Please choose the one that you think is the most important for you:

Over the long run, ESG portfolios will outperform the market.

```
ESG portfolios are more likely to hold their value – or increase in value – if climate
    risks materialize.
It's the right thing to do.
None of the above; there is no specific reason to invest in ESG portfolios.
```

Level of concern about climate change, and its five options in this order.

```
How would you describe your level of concern about climate change?

Extremely concerned
Very concerned
Somewhat concerned
Not very concerned
Not at all concerned
```

Delivery formats

The web form's input box becomes a rule for numeric answers that repeats the questionnaire's own answer note and unit label, and its radio buttons become lettered options with a single-choice rule.

Numeric answer rules, market question then ESG question.

```
Answer with a number only, in percent per year. (Please answer only with a positive or
    negative numeric value, with at most 1 decimal.) Your answer is read as "% per
    year, over the next 10 years".

Answer with a number only, in percent per year. (Please answer only with a positive or
    negative numeric value with at most 1 decimal.) Your answer is read as "% per
    year, over the next 10 years".
```

Single-choice rules, four options then five.

```
Choose exactly one option and answer with that option's capital letter only (A, B, C,
    or D).

Choose exactly one option and answer with that option's capital letter only (A, B, C,
    D, or E).
```

Validation identities

The validation exercises use identity templates distinct from those of the main round. In the first exercise of Section 7.1, each profile specifies age, gender, brokerage wealth, residential flood-risk exposure, and county political majority. Each attribute is drawn to match its marginal distribution in the published sample.

Identity template for the first exercise. [AGE] is an integer and [WEALTH] a one-decimal amount in thousands; [GENDER] is Male or Female, [FLOOD RISK] is High, Medium, or Low, and [POLITICAL VIEW] is Democratic or Republican.

```
You are a retail investor. You are [AGE] years old. Your gender is [GENDER]. Your
    wealth in the brokerage account is [WEALTH] thousand U.S. dollars. The flood risk
    exposure of your residential area is [FLOOD RISK]. Your local county's political
    majority, based on a recent U.S. presidential vote share among Democratic and
    Republican votes, is [POLITICAL VIEW].
```

SCF household description. Each description combines characteristics from one household's 2022 SCF record. [EQUITY CLAUSE] either states that the household has no measured stock-market equity exposure or lists its nonzero equity holdings. The replication package provides the complete descriptions, one per household.

SCF question X7556 asks for a self-assessment of personal-finance knowledge on a scale from zero to ten, but its codebook records the zero-response category as -1 (Board of Governors, 2023). The administered description used the stored code and described the scale as -1 to 10. We reproduce this wording below as administered, although it differs from the SCF response scale.

```
You are a U.S. household investor represented in the 2022 Survey of Consumer Finances.
    You are in the [AGE GROUP] age group, and your gender is [GENDER], and your race
    or ethnicity is [RACE], and your education level is [EDUCATION]. Your annual
    household income is $[INCOME], placing your household around the [PERCENTILE]
    percentile of the household income distribution. Your normal annual household
    income is $[NORMAL INCOME], around the [NORMAL PERCENTILE] percentile of the
    normal-income distribution. [EQUITY CLAUSE] Your household has net worth of $[NET
    WORTH], total assets of $[TOTAL ASSETS], financial assets of $[FINANCIAL ASSETS],
    and total debt of $[TOTAL DEBT]. Your monthly rent is $[RENT]. Your Buy Now, Pay
    Later status is: [BNPL STATUS]. Your financial risk attitude is that you are [RISK
    ATTITUDE]. Your self-assessed personal-finance knowledge is self-rated
```

```
personal-finance knowledge [K] on a -1 to 10 scale. You answered [C] out of 3
financial-literacy question(s) correctly. In a hypothetical financial emergency,
you would [EMERGENCY]. Use these characteristics as your background identity when
answering subsequent investment and financial-choice questions.
```

Knowledge-date instruction

The validation prompt specifies December 2023 as the reference date for the first exercise and 2022 for the SCF exercise. The human benchmark pools survey waves from June 2021 through December 2023. These dates are instructions in the prompts and do not restrict the models' actual training information.

```
Answer as if your knowledge cutoff were [YEAR OF KNOWLEDGE] and you were living at
that time. Use only information available on or before that date. If a question
would require newer information, answer with the view you would have held at that
time.
```

Validation respondent stance

The validation round uses a separate stance paragraph, similar but not identical to the main round's retail stance: it speaks of respondents rather than investors, limits the point-estimate rule to the numeric questions, and adds a rule for the multiple-choice items.

```
Answer as this person, not as an analyst, a financial adviser, a representative for
any group, or an average investor. Your answers should reflect your own views,
including any disagreement with expert or consensus opinion. Respondents who share
these characteristics do not all answer alike, so answer as this particular
respondent rather than reporting a category average. Give a point estimate for
each numeric question, and do not mechanically report round or focal numbers. For
each multiple-choice question, choose the option that best reflects your own
motivation or concern.
```

User message and assembly

The validation user message is identical across respondents, exercises, and models. It contains the task and answer-format instructions followed by the four questions, their response

instructions, and their options in the order shown above. Each question is introduced by its JSON key. The task and answer-format instructions are reproduced below:

```
You are taking a survey. Answer the four questions below as the respondent described
    in the instructions.

Return only valid JSON with exactly these four keys, each value a string.
{
  "question_stock": "...",
  "question_1": "...",
  "question_2": "...",
  "question_3": "..."
}
Do not return explanations, confidence scores, markdown, additional keys, or any text
    before or after the
JSON object.
```

System-message assembly. The identity description, the knowledge-date instruction, the stance paragraph, and the main round's output rule are joined by single spaces, in this order.

```
[IDENTITY DESCRIPTION]  (stripped)
    + " " + knowledge_cutoff_rule([YEAR OF KNOWLEDGE])
    + " " + RESPONDENT_PERSPECTIVE
    + " " + OUTPUT_RULE
```

A.4 Dated-Sample Instruments

The dated round uses the main-round questions unchanged. It replaces two system instructions: the economy-state description and the rule governing which information the model should use. The information card supplies the environment, so the code rejects a cell that also contains an economy-state description.

The archived requests of the dated round specify `gpt-5.6-sol`, high reasoning effort, and at most 16,000 output tokens. Temperature and `top_p` were not specified, so the provider's defaults applied, and their values are unknown. The model name that was requested and returned is an alias and does not identify a fixed version of the model. The replication

package records these settings, the provenance of the dated-round calls, the archived calls, and the metadata available for the other rounds (see `Data/README.md`).

The two replaced standing instructions

Card-context rule, replacing the economy-state description.

```
At the time you answer, the information environment you face is the one described in
    the information card included with this survey. Read the whole card before
    answering. Treat everything in it as simultaneously true and available to you at
    the time of the decision. Do not assume that these starting conditions must
    persist for all ten years.
```

Card-vintage rule, replacing "Answer using what you already know at the time you are asked." The instruction below asks the respondent to answer from the information card and its own values and judgment, without identifying a calendar year or adding outside historical knowledge.

```
Base your answer on the information card and on your own values and judgment as this
    respondent. The card's ten modules were selected separately for this study and the
    card as a whole does not describe any actual calendar year, so do not try to
    identify a year and do not use outside knowledge to add to, correct, or reconcile
    what the card says.
```

System-message assembly.

```
[identity sentences, 6 for retail / 5 for institutions]
    + CARD_ENVIRONMENT_RULE
    + RETAIL_PERSPECTIVE or INSTITUTION_PERSPECTIVE
    + MODULE_RULES[module]
    + CARD_VINTAGE_RULE
    + OUTPUT_RULE
```

User-message construction

The user message begins with a header and the information card, followed by a separator of 72 hyphens and the unchanged main-round question block. Complete requests with mixed

cards averaged about 30,000 input tokens, including the information card, the respondent instructions, and the survey questions.
Card header.

```
INFORMATION CARD FOR THIS SURVEY
The ten modules below describe the information available to you when you make this
    decision.
```

Composition of the user message.

```
user = f"{CARD_HEADER}\n{card}\n\n{'-' * 72}\n\n{base.user}"
```

Information-card structure

All 300 cards, mixed and single-year, use the same heading formats, module order, and overview lines, but their identifiers and module contents differ. The ten modules are organized into the three parts below, in an order that differs from the list in Section 8.1. The corporate environmental and social characteristics module (Module 8 below) is held fixed across cards, while the other nine modules vary by source year.
Card-formatting specification.

```
PUBLIC_TEXT_RENDERER_CONTRACT = {
    "renderer_contract_version": "public-text-renderer-v2",
    "document_title": "Information Provided for This Survey",
    "card_id_format": "Survey card ID: {card_id}",
    "instructions_heading": "Instructions",
    "part_heading_format": "Part {part_number}. {part_title}",
    "module_heading_format": "Module {section_number}. {section_title}",
    "overview_format": "Overview: {section_overview}",
}
```

Common card structure, illustrated with one card identifier.

```
Information Provided for This Survey
Survey card ID: CARD_ABSDCPMBCAFHFSZNUHWSDLUR
Instructions
Part 1. Economic and Real-World Conditions
Module 1. Economy and Financial Markets
```

```
Overview: This section describes broad economic conditions, interest rates, inflation,
    market volatility, and financing conditions.
Module 2. Climate, Disasters, and Physical Losses
Overview: This section describes physical climate conditions, natural disasters, and
    associated human and economic losses.
Module 3. Work, Safety, and Equal Treatment
Overview: This section describes workplace safety, labor conflict, equal-employment
    conditions, and related social outcomes.
Part 2. Beliefs, Information, and Rules
Module 4. Public Views and Social Attitudes
Overview: This section summarizes public attitudes toward environmental, social, and
    governance issues.
Module 5. News and Investor Discussion
Overview: This section summarizes the volume and tone of relevant news coverage and
    investor discussion.
Module 6. Public Policy and Regulation
Overview: This section describes public-policy and regulatory actions affecting
    environmental, social, governance, and investment decisions.
Module 7. Corporate Conduct and Enforcement
Overview: This section describes corporate conduct, accountability, enforcement, and
    governance-related cases.
Part 3. Firms and Investment Markets
Module 8. Corporate Environmental and Social Characteristics
Overview: This section describes firms' environmental and social characteristics,
    including emissions, operating practices, and differences between greener and
    browner firms.
Module 9. Sustainable Investment Market
Overview: This section describes the size, composition, and development of the
    sustainable-investment market.
Module 10. Ownership and Capital Allocation
Overview: This section describes who owns financial assets and how capital is
    allocated across investor groups.
```

The year-guessing probe

The year-guessing probe presents the ten information modules to the main model, specifies the source-year range of 2014–2025, and asks which single year the card most plausibly

describes. Before the modules are submitted, the code removes the card title, the identifier, and the introductory instructions, including the instruction not to identify a year. The same procedure is applied to the 288 mixed cards and the twelve single-year benchmark cards.
System message.

```
You are an experienced U.S. financial-markets analyst. You will see the ten
    information modules of a survey information card describing economic, market,
    policy, energy, climate, social, and corporate conditions. Explicit dates have
    been removed from the text. The underlying material is drawn from the period 2014
    to 2025. Your task is to infer which single calendar year in that range the card
    as a whole most plausibly describes. Use everything you know about that period.
    Return only the requested JSON object.
```

The question appended after the ten modules.

```
------------------------------------------------------------------------

Which single calendar year between 2014 and 2025 does this card as a whole most
    plausibly describe?
Return only valid JSON with exactly these keys: {"year": 0, "confidence_0_100": 0}.
    "year" must be an integer between 2014 and 2025. Do not add any other text.
```

Example information card

The example below reproduces the opening of one of the 288 mixed cards: its title, identifier, instructions, and first two modules. When the twelve single-year benchmark cards are used to elicit WTP and beliefs directly, they keep the same introductory instructions. The instructions state that the card describes no actual calendar year, although each benchmark card draws on a single source year. The year-guessing probe removes these introductory instructions, as described above.

The economy module in this example uses 2023 information, and the climate module uses 2016 information.

```
Information Provided for This Survey
Survey card ID: CARD_ABSDCPMBCAFHFSZNUHWSDLUR

Instructions
```

Please read all ten modules below before answering the survey question. The information was assembled for this study. The ten modules were selected separately and may come from different historical settings, so the full card does not describe any actual calendar year. For this survey, assume that all information in the ten modules is simultaneously true and was available when you made the investment decision. Answer as the investor described in the respondent profile supplied with the survey. Do not try to identify a year. Do not use outside historical knowledge to add, correct, or reconcile any detail. Base your answer only on the respondent profile, this card, and the survey question.

Part 1. Economic and Real-World Conditions

Module 1. Economy and Financial Markets
Overview: This section describes broad economic conditions, interest rates, inflation, market volatility, and financing conditions.

Economic and market conditions
- 10-year U.S. Treasury yield at period end: 3.88%
- 10-year minus 3-month U.S. Treasury yield spread at period end: -1.52 percentage points
- CBOE market volatility index at period end: 12.45
- Change in the WTI crude-oil spot price over the information period: -10.32%
- Consumer price inflation (latest year-over-year reading available by period end): 3.14%

Economic conditions and monetary policy

Background summary 1
- Recent indicators suggest that growth of economic activity has slowed from its strong pace in the third quarter.
- Job gains have moderated since earlier in the year but remain strong, and the unemployment rate has remained low.
- Tighter financial and credit conditions for households and businesses are likely to weigh on economic activity, hiring, and inflation.
- The Committee remains highly attentive to inflation risks.
- The Committee seeks to achieve maximum employment and inflation at the rate of 2 percent over the longer run.

- In support of these goals, the Committee decided to maintain the target range for the federal funds rate at 5-1/4 to 5-1/2 percent.
- In determining the extent of any additional policy firming that may be appropriate to return inflation to 2 percent over time, the Committee will take into account the cumulative tightening of monetary policy, the lags with which monetary policy affects economic activity and inflation, and economic and financial developments.
- The Committee is strongly committed to returning inflation to its 2 percent objective.
- The Committee's assessments will take into account a wide range of information, including readings on labor market conditions, inflation pressures and inflation expectations, and financial and international developments.

Module 2. Climate, Disasters, and Physical Losses
Overview: This section describes physical climate conditions, natural disasters, and associated human and economic losses.

Federally declared disasters
- Federal disaster declarations for biological incidents: 0
- Federal disaster declarations for chemical incidents: 0
- Federal disaster declarations for coastal storms: 0
- Federal disaster declarations for dam or levee failures: 0
- Federal disaster declarations for earthquakes: 0
- Federal disaster declarations for fires: 51
- Federal disaster declarations for floods: 18
- Federal disaster declarations for hurricanes: 10
- Federal disaster declarations for mudslides or landslides: 1
- Federal disaster declarations for severe ice storms: 0
- Federal disaster declarations for severe storms: 16
- Federal disaster declarations for snowstorms: 5
- Federal disaster declarations for straight-line winds: 0
- Federal disaster declarations for tornadoes: 1
- Federal disaster declarations for toxic-substance incidents: 1
- Federal disaster declarations for tropical depressions: 0
- Federal disaster declarations for tropical storms: 0
- Federal disaster declarations for typhoons: 0
- Federal disaster declarations for volcanic eruptions: 0
- Federal disaster declarations for winter storms: 0
- Federal disaster declarations in other incident categories: 0

```
- Total federal disaster declarations: 103

#### Atlantic tropical-storm activity
- Hurricanes in the Atlantic basin: 7
- Major hurricanes in the Atlantic basin: 4
- Named storms in the Atlantic basin: 15

#### Economic and insured catastrophe losses
- Economic losses from human-caused catastrophes: Not reported
- Economic losses from natural catastrophes: $150 billion in inflation-adjusted U.S.
    dollars at the source's published price basis
- Insured losses from human-caused catastrophes: $7 billion in inflation-adjusted U.S.
    dollars at the source's published price basis
- Insured losses from natural catastrophes: $42 billion in inflation-adjusted U.S.
    dollars at the source's published price basis
- Total economic losses from catastrophes: $158 billion in inflation-adjusted U.S.
    dollars at the source's published price basis
- Total insured losses from catastrophes: $49 billion in inflation-adjusted U.S.
    dollars at the source's published price basis
```

The excerpt omits the remaining eight modules, which contain approximately 1,000 additional lines. The replication package provides the complete text of all 288 mixed cards and twelve single-year benchmark cards.

B Additional Exhibits

This appendix reports additional exhibits for Sections 4 through 7 and for the time series of Section 8. For the time series, the tables give the design of the dated sample, the sources of the real-world series, and the sensitivity of their regressions to sentiment controls; the figures give the reconstructed series by political affiliation and by investor class, the comparison of reconstructed with direct single-year answers, and the placebo distribution of the module R^2.

Table B.1. The two margins and retail attributes

	Share of wealth in stocks		Share of stocks in Portfolio H	
	Coefficient	(SE)	Coefficient	(SE)
Political affiliation (omitted: Neutral)	$\Delta R^2 = 0.007$		$\Delta R^2 = 0.931$	
Strong Democrat	2.45^{***}	(0.54)	16.17^{***}	(0.45)
Weak Democrat	0.53	(0.54)	6.90^{***}	(0.41)
Weak Republican	0.86	(0.57)	-25.19^{***}	(0.58)
Strong Republican	3.00^{***}	(0.53)	-39.36^{***}	(0.47)
Income (omitted: 40th to 80th percentile)	$\Delta R^2 = 0.475$		$\Delta R^2 = 0.004$	
Bottom 40%	-16.70^{***}	(0.52)	-2.58^{***}	(0.43)
80th to 95th percentile	5.56^{***}	(0.38)	0.53	(0.45)
Top 5%	8.27^{***}	(0.43)	0.75	(0.46)
Age (omitted: 45 to 64)	$\Delta R^2 = 0.319$		$\Delta R^2 = 0.006$	
Under 35	10.84^{***}	(0.53)	3.40^{***}	(0.45)
35 to 44	4.98^{***}	(0.47)	1.58^{***}	(0.43)
65 or older	-10.79^{***}	(0.39)	-0.83^{**}	(0.42)
Financial literacy (omitted: medium)	$\Delta R^2 = 0.120$		$\Delta R^2 = 0.019$	
Low	-7.28^{***}	(0.40)	-4.57^{***}	(0.36)
High	4.56^{***}	(0.37)	2.65^{***}	(0.41)
Gender (omitted: male)	$\Delta R^2 = 0.012$		$\Delta R^2 = 0.015$	
Female	-3.08^{***}	(0.34)	5.18^{***}	(0.32)
Constant (reference identity)	52.67^{***}	(0.56)	65.80^{***}	(0.52)
R^2		0.933		0.975
Identity cells		480		480

Notes: Each column reports a regression of one stated allocation share from ChatGPT-5.6 Sol on category indicators for all five attributes at once: politics, income, gender, financial literacy, and age. The sample contains 480 baseline retail cells, with three repeats averaged within each cell. The outcomes, in percent, are the share of financial wealth in U.S. stocks and the share of that stock allocation in Portfolio H. H and L are the only equity choices, are otherwise identical, and score at the 80th and 20th ESG percentiles; the economy is normal. The constant represents a politically Neutral male aged 45–64, with income in the 40th–80th percentile and medium financial literacy. Standard errors in parentheses are heteroskedasticity-robust and summarize variation in the model's answers across the designed identity grid. ***, **, and * denote $|t| \geq 2.58$, 1.96, and 1.645, respectively. ΔR^2 is the decline in R^2 when an attribute's indicators are removed from the full model.

Table B.2. The two margins and institutional attributes

	Share of AUM in U.S. equities		Share of equities in Portfolio H	
	Coefficient	(SE)	Coefficient	(SE)
Institution type (omitted: pension fund)	$\Delta R^2 = 0.738$		$\Delta R^2 = 0.546$	
Endowment or foundation	-0.43	(1.32)	3.39^{***}	(1.19)
Insurance company	-12.35^{***}	(1.36)	-3.37^{***}	(1.28)
Hedge fund	15.56^{***}	(1.39)	-13.28^{***}	(1.30)
Mutual fund company	11.99^{***}	(2.06)	-13.56^{***}	(1.22)
Geography (omitted: U.S.)	$\Delta R^2 = 0.113$		$\Delta R^2 = 0.378$	
Europe	-9.30^{***}	(2.07)	15.87^{***}	(0.97)
Asia	-8.90^{***}	(2.04)	5.54^{***}	(0.89)
Other regions	-8.68^{***}	(2.12)	5.22^{***}	(1.17)
AUM size (omitted: medium)	$\Delta R^2 = 0.020$		$\Delta R^2 = 0.003$	
Small	1.54	(1.49)	-0.79	(0.79)
Large	-2.40^{**}	(1.15)	0.47	(0.85)
Constant (reference institution)	27.44^{***}	(1.47)	69.17^{***}	(1.00)
R^2		0.871		0.927
Identity cells		60		60

Notes: Each column reports a regression of one stated allocation share from ChatGPT-5.6 Sol on category indicators for all three attributes at once: institution type, geography, and assets under management (AUM). The sample contains 60 baseline institutional cells, with three repeats averaged within each cell. The outcomes, in percent, are the share of AUM in U.S. public equities and the share of that equity allocation in Portfolio H. H and L are the only equity choices, are otherwise identical, and score at the 80th and 20th ESG percentiles; the economy is normal. The constant represents a medium-size U.S. pension fund. Standard errors in parentheses are heteroskedasticity-robust and summarize variation in the model's answers across the designed identity grid. ***, **, and * denote $|t| \geq 2.58$, 1.96, and 1.645, respectively. ΔR^2 is the decline in R^2 when an attribute's indicators are removed from the full model.

Table B.3. Portfolio H share, WTP, and the matched belief gap by investor class and with an institution indicator

	ChatGPT-5.6 Sol			ChatGPT-5.5			Claude Opus 5.0			Claude Fable 5		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Retail cells (2,880)												
WTP	43.95***		32.39***	49.53***		36.67***	21.99***		17.89***	40.78***		34.79***
	(0.96)		(1.16)	(1.02)		(1.11)	(0.15)		(0.32)	(0.35)		(0.70)
Matched belief gap		27.38***	11.77***		23.71***	10.95***		12.49***	2.82***		13.56***	2.45***
		(0.43)	(0.57)		(0.41)	(0.43)		(0.11)	(0.19)		(0.16)	(0.26)
R^2	0.61	0.48	0.66	0.62	0.48	0.68	0.89	0.77	0.90	0.85	0.72	0.85
Panel B: Institutional cells (360)												
WTP	77.35***		77.72***	71.56***		71.28***	35.58***		34.84***	65.06***		51.86***
	(3.89)		(3.92)	(4.21)		(4.23)	(2.39)		(2.61)	(2.18)		(2.44)
Matched belief gap		−9.22	6.57		15.33*	10.32*		8.06***	0.98		23.51***	10.05***
		(10.31)	(6.24)		(8.71)	(6.04)		(1.07)	(0.93)		(1.17)	(1.06)
R^2	0.58	0.00	0.58	0.45	0.01	0.45	0.62	0.12	0.62	0.71	0.47	0.77
Panel C: All cells (3,240), with an institution indicator												
WTP	44.22***		33.09***	49.81***		37.40***	22.23***		18.82***	41.12***		35.27***
	(0.96)		(1.15)	(1.02)		(1.11)	(0.15)		(0.32)	(0.35)		(0.68)
Matched belief gap		27.34***	11.42***		23.70***	10.70***		12.38***	2.37***		13.65***	2.41***
		(0.43)	(0.56)		(0.41)	(0.43)		(0.12)	(0.20)		(0.16)	(0.25)
Institution indicator	20.97***	11.76***	19.21***	25.59***	17.53***	23.94***	11.86***	10.02***	11.67***	14.20***	13.28***	14.23***
	(0.46)	(0.54)	(0.48)	(0.53)	(0.65)	(0.54)	(0.35)	(0.49)	(0.36)	(0.40)	(0.51)	(0.40)
R^2	0.62	0.48	0.66	0.63	0.49	0.69	0.87	0.74	0.88	0.84	0.71	0.84

Notes: The table reestimates the regressions of Table 5 within each investor class (Panels A and B) and over all cells with an indicator equal to one for institutional identities (Panel C). The dependent variable is the stated share of the equity allocation assigned to Portfolio H, in percent. For each model, the three columns use willingness to pay (WTP), the matched belief gap, and both, with variables and units as in Table 5 and an intercept that is not shown. In the third column of each model in Panel C, the indicator's coefficient is the difference in the Portfolio H share between an institutional and a retail identity with the same WTP and matched belief gap. The indicator alone gives 8.4, 15.8, 7.6, and 8.5 percentage points over the same cells. Standard errors in parentheses are heteroskedasticity-robust, as in Table 5, and summarize variation across the designed identity grid. With standard errors clustered by identity, all significance levels are unchanged except for the matched belief gap for ChatGPT-5.5 in Panel B, columns (5) and (6), which is significant at the 10% level with robust standard errors and not with clustered ones. ***, **, and * denote $|t| \geq 2.58$, 1.96, and 1.645, respectively.

Table B.4. Own choice versus outside pressure

	ChatGPT-5.6 Sol	ChatGPT-5.5	Claude Opus 5.0	Claude Fable 5
Retail: own-choice share	92.6 (0.19)	91.2 (0.21)	83.9 (0.20)	81.4 (0.43)
Institutions: own-choice share	41.4 (1.57)	52.0 (1.82)	41.9 (1.35)	50.2 (1.58)

Notes: The table reports the mean percentage of the stated Portfolio H holding attributed to own choice across 480 retail and 60 institutional baseline identity cells. The baseline uses aggregate ESG scores and a normal economy. Repeats are averaged within each cell, with standard errors across cells in parentheses. Portfolio H is the higher-scoring of two otherwise identical portfolios at the 80th and 20th percentiles of the same greenness measure. Each identity divides its H holding between own choice and outside pressure, with the shares summing to 100. For retail identities, outside pressure comes from friends, family, colleagues, or others whose opinion the investor cares about. For institutions, it includes mandates, the investor base, trustees or clients, regulatory or industry expectations, and reputational considerations. Appendix A reproduces both questions and pressure lists.

Table B.5. The 13 models and cutoffs from current provider documentation

Model	Provider	Documented cutoff	Role in the paper	Cells answered
GPT-4o	OpenAI	October 2023	Further version	240
GPT-4o mini	OpenAI	October 2023	Further version	240
GPT-4 Turbo	OpenAI	December 2023	Further version	240
GPT-4.1	OpenAI	June 2024	Further version	240
GPT-5.4	OpenAI	August 2025	Further version	240
GPT-5.4 mini	OpenAI	August 2025	Further version	240
ChatGPT-5.5	OpenAI	December 2025	Full-grid robustness	3,240
ChatGPT-5.6 Sol	OpenAI	February 2026	Main model	3,240
Claude Haiku 4.5	Anthropic	February 2025	Further version	240
Claude Fable 5	Anthropic	January 2026	Full-grid robustness	3,240
Claude Opus 4.8	Anthropic	January 2026	Further version	240
Claude Sonnet 5	Anthropic	January 2026	Further version	480
Claude Opus 5.0	Anthropic	May 2026	Full-grid robustness	3,240

Notes: The table lists thirteen models, ordered by provider and then by the cutoff month in provider documentation accessed September 5, 2026. Dates come from OpenAI's model documentation and Anthropic's Transparency Hub and are displayed to the month. The exact URLs and fields are in `Data/Provenance/`. The accessed OpenAI documentation gives February 16, 2026 for ChatGPT-5.6 Sol. Our archived configuration records a May 2026 cutoff for the same model. No official source for that date has been found, and the archived configuration is kept as recorded. Request and return records give each model's name but cannot verify its training cutoff or tie the answers to a fixed version of the model. Cells answered refers to coverage of the 3,240 identity-condition cells in the main round. Each full-grid model answers all cells; Claude Sonnet 5 answers a shared 480-cell subsample, and the remaining eight versions answer a 240-cell subsample nested within it. The version figures compare all thirteen models on those same 240 cells. The nine further versions answer only the preference module and the two belief modules, for a total of 7,200 calls (Table 1).

Table B.6. Dated-sample design summary

Design element	Count
Historical information environments (years, 2014–2025)	12
Information modules per card (nine year-varying, one fixed)	10
Mixed cards (at least four distinct source years among the nine year-varying modules, at most three modules from any single year)	288
Cards carrying each module–year pair	24
Identities (five political affiliations at a fixed retail profile; two institution types)	7
Question modules per card (preference, matched beliefs, unrestricted beliefs)	3
Main calls (7 identities × 3 questions × 288 cards)	6,048
Repeat calls (one identity; 60 cards asked twice more per question)	360
Single-year benchmark calls (7 × 3 × 12 cards)	252
Year-guessing calls (288 mixed and 12 single-year cards, one each)	300
Total calls	6,960

Notes: All calls use ChatGPT-5.6 Sol. Appendix A.4 reports the request settings. The five retail political identities share a fixed profile: age 45–64, income in the 40th–80th percentile, male, and high financial literacy. All 6,960 calls returned usable answers.

Table B.7. Sources and definitions of the real-world series

Series	Source and definition
Flows: net flows into U.S. sustainable funds	Annual net flows in USD billion for 2014–2024 are digitized from the chart on p. 10 of Morningstar's *U.S. Sustainable Funds Landscape: 2024 in Review* (January 2025), using the axis ticks. The 2023 and 2024 readings are −13.29 and −19.54, compared with published text values of −13.3 and −19.6. The same method applied to the February 2023 edition reproduces its published 2017–2019 figures. Following a definition change after August 2022, Morningstar restated 2019 flows from 21.4 to 15.8. The series uses the January 2025 vintage throughout.
Fee: ESG fee of Baker, Egan, and Sarkar	Baker et al. (2024), Table 2, column (2): the additional annual fee, in basis points, that investors are willing to pay for an index fund with an ESG mandate, estimated from index-fund flows. Their table reports one estimate per year for 2019–2023, with 2019 covering May–December.
Launches: sustainable-fund launches	Number of sustainable funds launched in each year, active and passive, including funds since liquidated, from the chart on p. 24 of the January 2025 Morningstar edition (2015–2024); the 2025 count is from the text of Morningstar's March 2026 review of 2025. The observation for year t is the count for $t+1$.
Concern: Media Climate Change Concerns index	Ardia et al. (2023): articles on climate change in major U.S. newspapers and newswires, scored for risk language and negative tone and aggregated by volume; the monthly aggregate index (sentometrics-research.com) averaged by year, with 2025 the January–June mean.
Google: search interest in ESG	Google Trends, United States, query "ESG," monthly relative search volume over the single window January 2013–December 2025 (peak = 100), averaged by year.
Calls: earnings-call climate exposure	Sautner et al. (2023): share of climate-related bigrams in earnings-call transcripts; the firm-year measure, equal-weighted across U.S.-headquartered firms, in units of one per thousand, 2014–2024.
Sentiment controls	Three measures not specific to climate or ESG: the Federal Reserve Bank of San Francisco daily news sentiment index (Shapiro et al., 2022), the University of Michigan index of consumer sentiment, and the CBOE VIX, each averaged by year.

Notes: The reconstructed willingness to pay (WTP) and unrestricted belief gap are the weighted ChatGPT-5.6 Sol series in Figure 9. Every real-world series is annual. Subannual index, search-volume, and sentiment observations are averaged within calendar years as specified in the table; net flows, fund-launch counts, and earnings-call exposure retain the annual definitions in their rows. Section 8.7 relates these series to the reconstructed tastes and beliefs.

Table B.8. Real-world series with sentiment controls: investing series

Regressor	(1) Baseline	(2) + News sentiment	(3) + Consumer sentiment	(4) + VIX	(5) + All three
Flows (Morningstar)					
WTP	0.78***	0.74***	0.77**	0.69**	0.68**
	(3.46)	(3.56)	(3.22)	(3.39)	(3.03)
Beliefs	0.02	0.06	0.19	0.09	−0.28
	(0.11)	(0.27)	(0.47)	(0.46)	(−0.63)
News sentiment		−0.33			−0.10
		(−1.65)			(−0.23)
Consumer sentiment			−0.20		0.48
			(−0.51)		(0.94)
VIX				0.37	0.48
				(1.88)	(1.03)
Observations / R^2	11 / 0.62	11 / 0.73	11 / 0.64	11 / 0.75	11 / 0.79
Fee (Baker, Egan, and Sarkar)					
WTP	0.70*	—	—	—	—
	(4.10)				
Beliefs	0.44	—	—	—	—
	(2.57)				
With five annual observations, adding a control leaves fewer than two residual degrees of freedom.					
Observations / R^2	5 / 0.95	—	—	—	—
Launches (Morningstar)					
WTP	0.74**	0.66***	0.70**	0.59***	0.59***
	(3.15)	(4.44)	(3.10)	(4.64)	(4.66)
Beliefs	0.11	0.15	0.51	0.21	−0.11
	(0.45)	(1.04)	(1.32)	(1.69)	(−0.44)
News sentiment		−0.52***			−0.16
		(−3.61)			(−0.66)
Consumer sentiment			−0.49		0.41
			(−1.29)		(1.42)
VIX				0.57***	0.60*
				(4.65)	(2.27)
Observations / R^2	11 / 0.60	11 / 0.86	11 / 0.67	11 / 0.90	11 / 0.93

Notes: The panels report annual regressions for the investing series in their headers, defined as in Table 8: Flows and next-year Launches over 2014–2024, and Fee over 2019–2023. Column (1) includes the reconstructed willingness to pay (WTP) and the reconstructed unrestricted belief gap (Beliefs) from ChatGPT-5.6 Sol; columns (2)–(4) add one sentiment control and column (5) all three. WTP averages the two annual-return tradeoffs for the matched portfolios; Beliefs is the expected return on typical real-world high-ESG portfolios minus that on low-ESG portfolios. All variables have within-sample mean zero and standard deviation one. Parentheses report OLS t-statistics; ***, **, and * denote two-sided $p < 0.01$, 0.05, and 0.10, using $n - k - 1$ degrees of freedom (k regressors, excluding the intercept). Specifications with fewer than two residual degrees of freedom are omitted (—). Controls are three general measures of sentiment: annual means of the Federal Reserve Bank of San Francisco daily news sentiment index (Shapiro et al., 2022), University of Michigan consumer sentiment, and CBOE VIX.

Table B.9. Real-world series with sentiment controls: attention series

Regressor	(1) Baseline	(2) + News sentiment	(3) + Consumer sentiment	(4) + VIX	(5) + All three
Concern (Ardia et al.)					
WTP	0.21	0.19	0.18	0.19	0.20
	(1.27)	(1.11)	(1.01)	(1.08)	(1.02)
Beliefs	−0.88***	−0.87***	−0.72*	−0.87***	−0.77
	(−5.37)	(−5.16)	(−2.25)	(−4.97)	(−1.90)
News sentiment		−0.14			−0.30
		(−0.81)			(−0.80)
Consumer sentiment			−0.19		−0.14
			(−0.61)		(−0.32)
VIX				0.07	−0.25
				(0.38)	(−0.63)
Observations / R^2	12 / 0.76	12 / 0.78	12 / 0.77	12 / 0.77	12 / 0.80
Google (Google Trends)					
WTP	0.06	0.01	−0.01	−0.01	−0.01
	(0.44)	(0.13)	(−0.08)	(−0.09)	(−0.06)
Beliefs	−0.91***	−0.88***	−0.50*	−0.86***	−0.73**
	(−6.37)	(−9.10)	(−2.21)	(−7.90)	(−3.18)
News sentiment		−0.33***			−0.26
		(−3.46)			(−1.22)
Consumer sentiment			−0.48*		−0.18
			(−2.16)		(−0.70)
VIX				0.31**	0.02
				(2.82)	(0.08)
Observations / R^2	12 / 0.82	12 / 0.93	12 / 0.89	12 / 0.91	12 / 0.93
Calls (Sautner et al.)					
WTP	0.46***	0.43***	0.43***	0.39***	0.39**
	(3.74)	(4.12)	(4.36)	(4.44)	(3.89)
Beliefs	−0.96***	−0.94***	−0.64***	−0.91***	−0.79**
	(−7.91)	(−9.18)	(−3.79)	(−10.57)	(−3.95)
News sentiment		−0.21*			0.07
		(−2.08)			(0.35)
Consumer sentiment			−0.38*		−0.15
			(−2.34)		(−0.67)
VIX				0.26**	0.25
				(3.06)	(1.21)
Observations / R^2	11 / 0.89	11 / 0.93	11 / 0.94	11 / 0.95	11 / 0.96

Notes: The panels report annual regressions for the attention series in their headers, defined as in Table 8: Concern and Google over 2014–2025 and Calls over 2014–2024. Column (1) includes the reconstructed willingness to pay (WTP) and the reconstructed unrestricted belief gap (Beliefs) from ChatGPT-5.6 Sol; columns (2)–(4) add one sentiment control and column (5) all three. Parentheses report OLS t-statistics; ***, **, and * denote two-sided $p < 0.01$, 0.05, and 0.10, using $n - k - 1$ degrees of freedom (k regressors, excluding the intercept). Standardization, variable definitions, and controls are as in Table B.8.

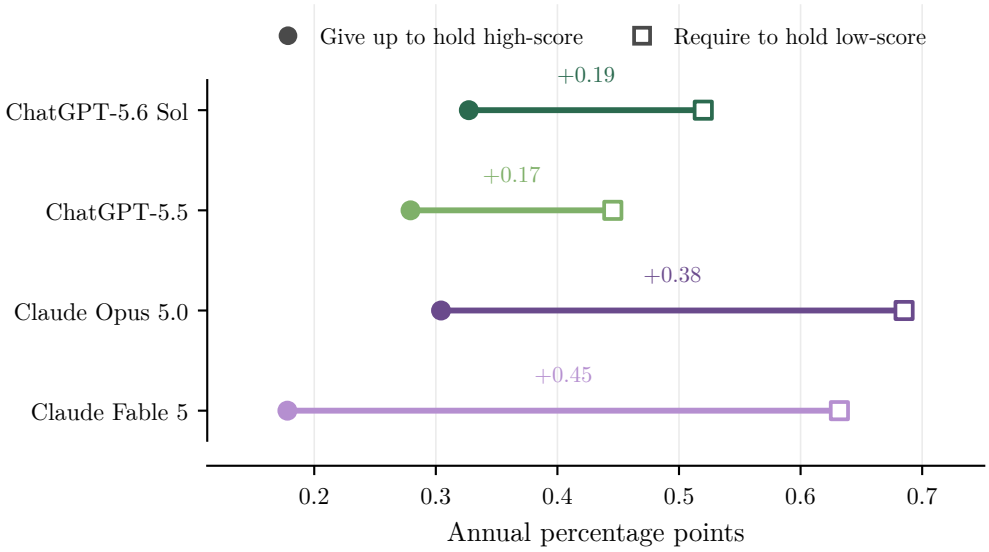


Figure B.1. The two sides of WTP across models. This figure plots the two components of willingness to pay (WTP) in the four models that answer the full grid, in annual percentage points. Circles show the annual return given up to hold Portfolio H; squares show the extra annual return required to hold Portfolio L. Portfolios H and L are otherwise identical and score at the 80th and 20th percentiles of the same greenness measure. WTP is the average of these two signed tradeoffs. Each row pools every answer across all 3,240 cells of the indicated model, including retail and institutional identities under all score and economy conditions. The three repeats per cell from ChatGPT-5.6 Sol are included. The tradeoffs refer to annual returns over ten years. The printed value is the extra return required for L minus the return given up for H.

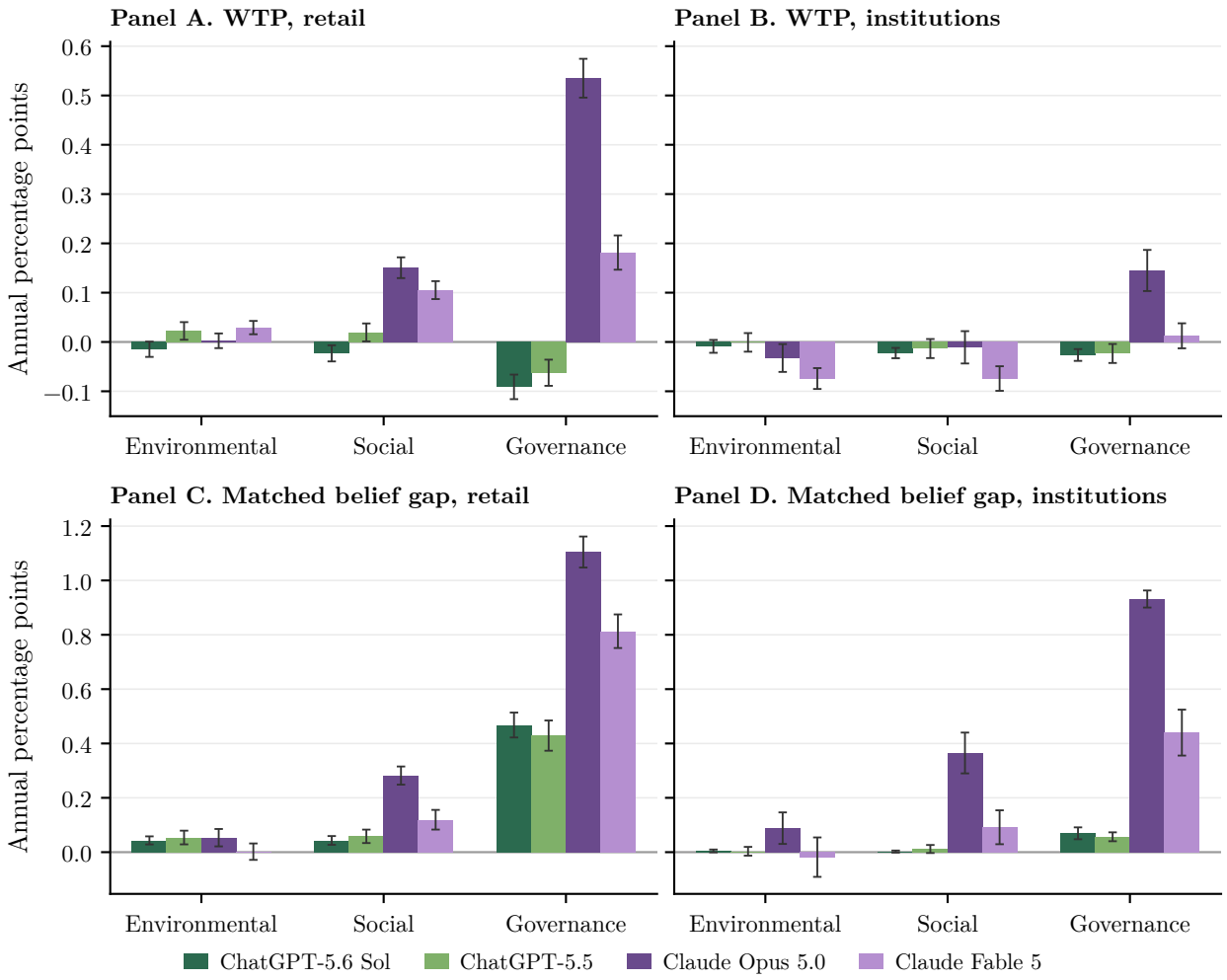


Figure B.2. Changes in tastes and beliefs across ESG dimensions by investor class. This figure reports within-identity changes in willingness to pay (WTP) and in the matched belief gap when the score measures environmental, social, or governance characteristics separately instead of aggregate ESG, in the four models that answer the full grid. WTP averages the annual return given up for Portfolio H and the extra annual return required for Portfolio L over ten years, and the matched belief gap is the expected annual return over ten years on H minus that on L. Both are in annual percentage points. H and L are otherwise identical and score at the 80th and 20th percentiles of the same measure. The economy is normal. Each difference subtracts the identity's aggregate-ESG baseline response, as in Figure 6. Panels A and B report WTP and Panels C and D the matched belief gap. Panels A and C show the 480 retail identities (480 pairs per bar), and Panels B and D the 60 institutional identities (60 pairs per bar). Panels in the same row share a vertical scale. Error bars are 95% confidence intervals, calculated as 1.96 standard errors of the paired differences.

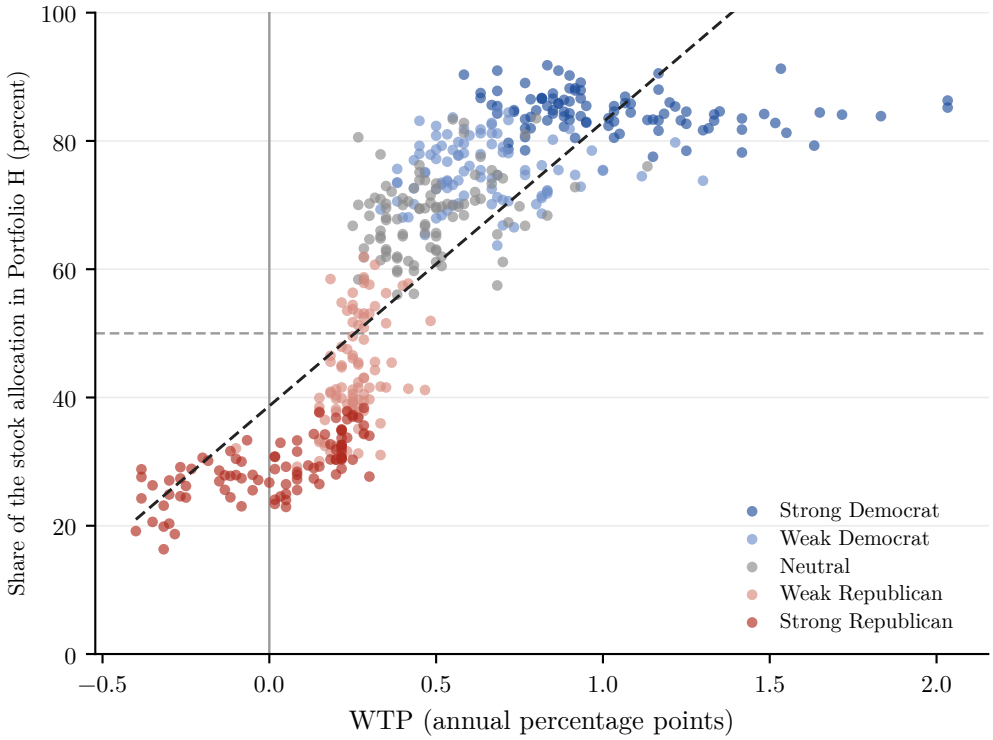


Figure B.3. Portfolio H share and WTP. This figure plots the stated share of the stock allocation assigned to Portfolio H, in percent, against willingness to pay (WTP), in annual percentage points, for the 480 baseline retail identities in ChatGPT-5.6 Sol. WTP is the average of the annual return given up for Portfolio H and the extra annual return required for Portfolio L over ten years. The portfolios are otherwise identical and score at the 80th (H) and 20th (L) percentiles of the same greenness measure. The baseline uses aggregate ESG scores and a normal economy. The allocation question offers H and L as the only equity choices. Each point averages the three repeats within an identity cell and is colored by political affiliation. The dashed black line is the least-squares fit. The horizontal line at 50 marks an equal split between H and L, and the vertical line marks zero WTP.

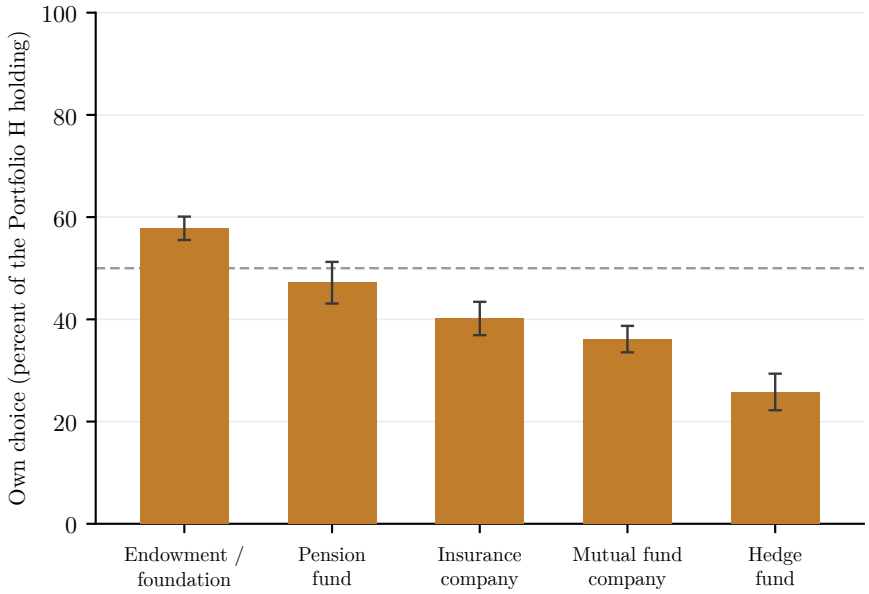


Figure B.4. Own choice by institution type. This figure plots the percentage of the stated Portfolio H holding attributed to own choice by institution type in ChatGPT-5.6 Sol. Portfolios H and L are otherwise identical and score at the 80th and 20th percentiles of the same greenness measure. The baseline uses aggregate ESG scores and a normal economy. Each institutional identity divides its H holding between own choice and outside pressure, with the two shares summing to 100. Each bar averages the 12 baseline cells of one institution type, after averaging the three repeats within each cell. The bars are ordered from the highest to the lowest mean. The dashed line at 50 marks equal shares for own choice and outside pressure. Error bars are 95% confidence intervals, calculated as 1.96 standard errors across cells.

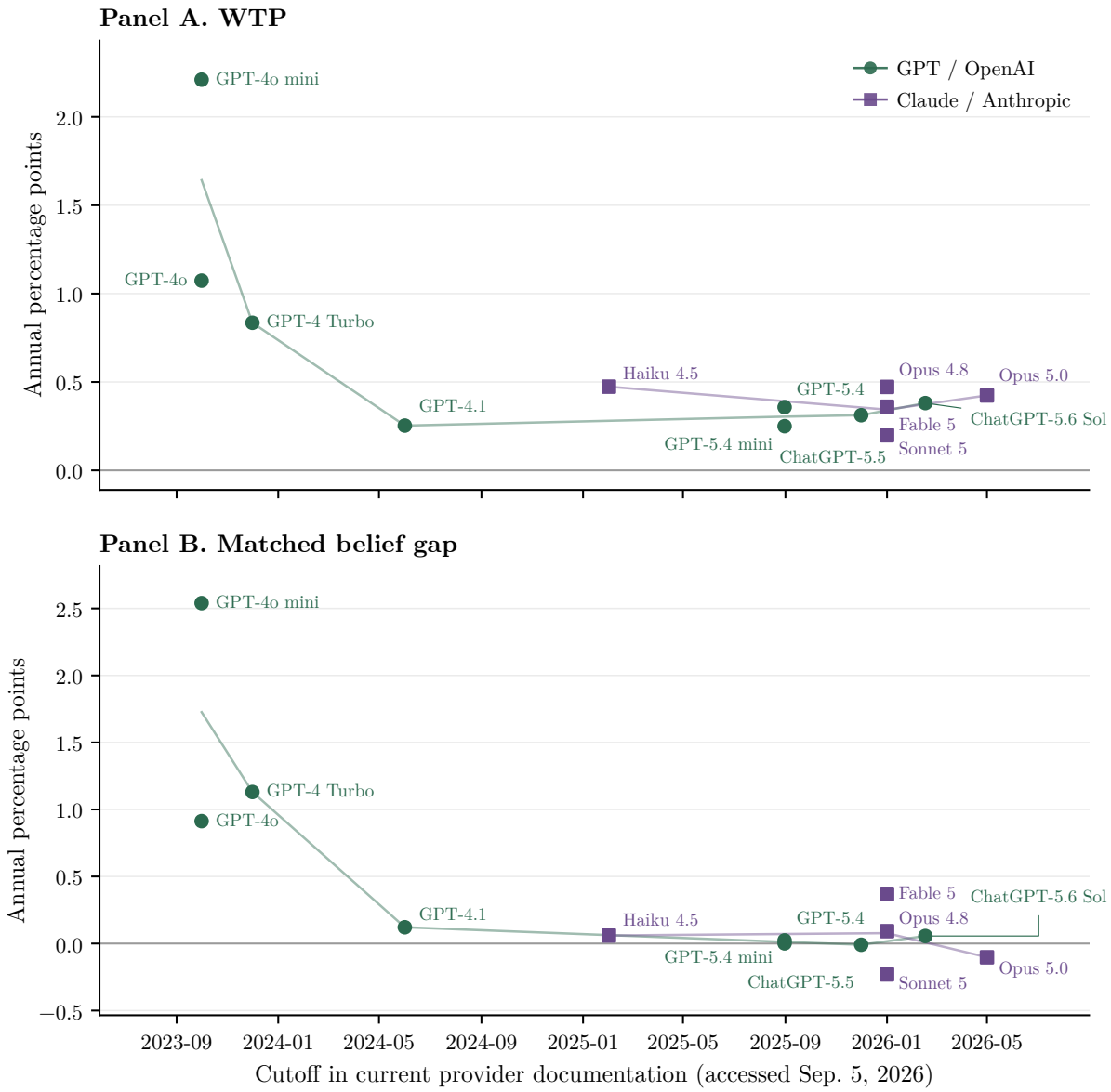


Figure B.5. Tastes and beliefs across 13 model versions. This figure compares willingness to pay (WTP; Panel A) and matched belief gaps (Panel B) across 13 models on the same 240 cells, in annual percentage points. WTP averages the return given up for Portfolio H and the extra return required for Portfolio L, and the gap is the expected return on H minus that on L. Both refer to annual returns over ten years on otherwise identical portfolios at the 80th and 20th score percentiles. The fixed subsample spans retail and institutional identities under all score and economy conditions. Models are placed at the cutoff dates given in provider documentation, accessed on September 5, 2026. Thin lines connect provider-family means at each date. Table B.5 gives full model names and sources.

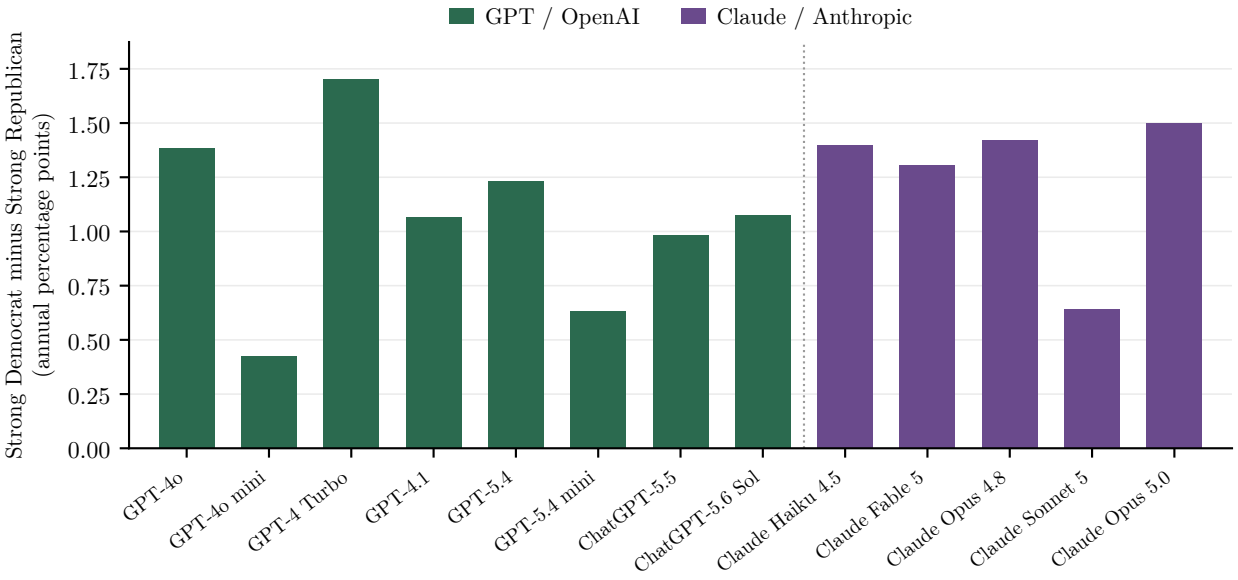


Figure B.6. The political WTP gap across 13 model versions. This figure plots the strong-Democrat-minus-strong-Republican difference in willingness to pay (WTP) across 13 models, in annual percentage points, using 75 baseline retail cells per model. WTP averages the annual return given up for Portfolio H and the extra annual return required for Portfolio L over ten years. H and L are otherwise identical and score at the 80th and 20th ESG percentiles; the economy is normal. Models are ordered within provider families by the cutoff dates in provider documentation. Table B.5 gives the sources of these dates and explains why they do not identify a fixed version of each model.

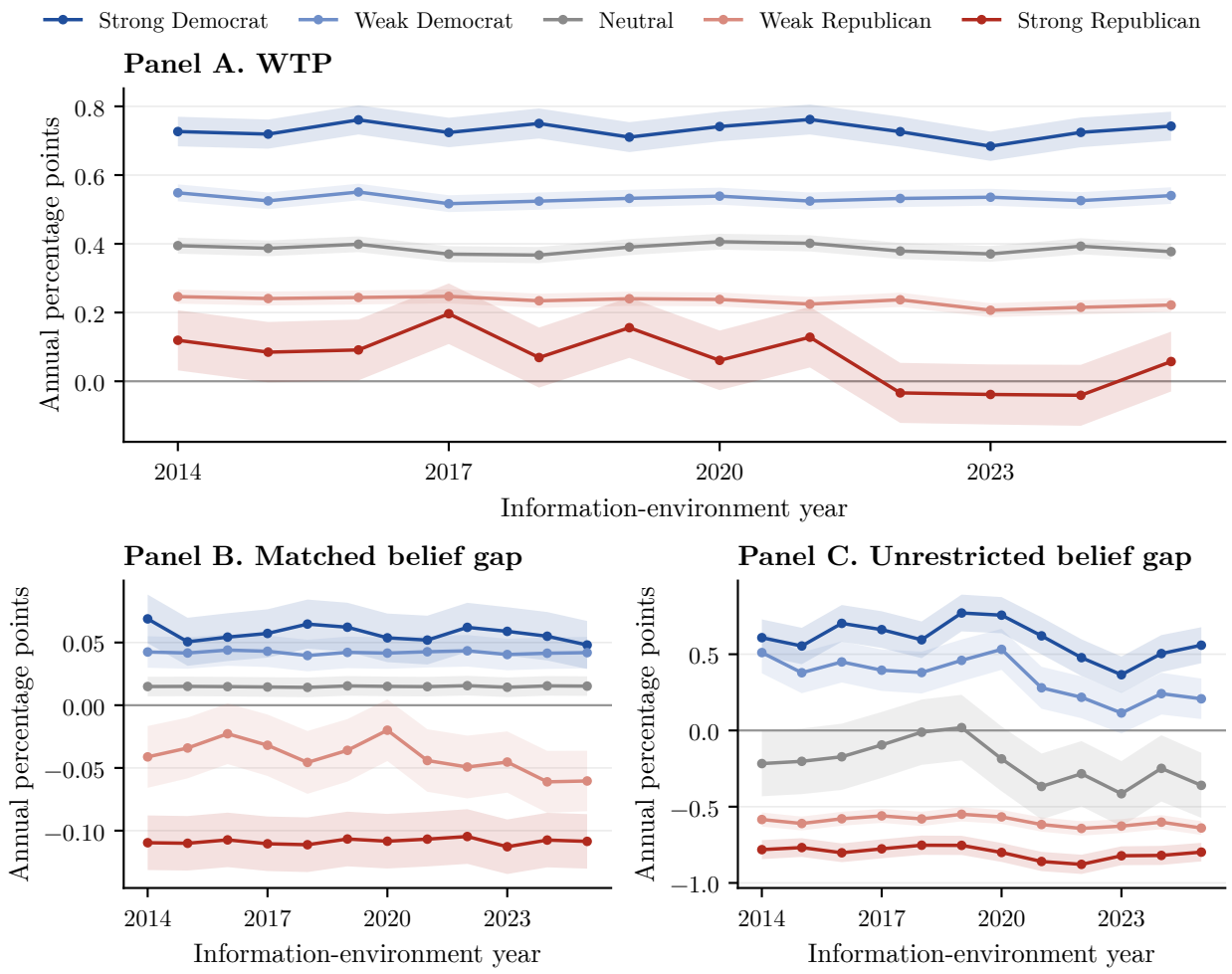


Figure B.7. Reconstructed series by political affiliation. This figure plots reconstructed willingness to pay (WTP) and belief gaps for five political identities in ChatGPT-5.6 Sol across the 2014–2025 information environments, in annual percentage points. Panel A reports WTP, Panel B the matched belief gap, and Panel C the unrestricted belief gap. WTP averages the annual return given up for Portfolio H and the extra return required for Portfolio L; belief gaps are the expected return on H minus that on L. Both refer to annual returns over ten years. H and L score at the 80th and 20th ESG percentiles and are otherwise identical in the preference and matched-belief questions; unrestricted beliefs refer to typical real-world portfolios. Each identity answers all 288 mixed cards. The other retail attributes are fixed at age 45–64, middle income, male, and high financial literacy. Shaded bands are 95% confidence intervals, calculated as 1.96 reconstruction standard errors.

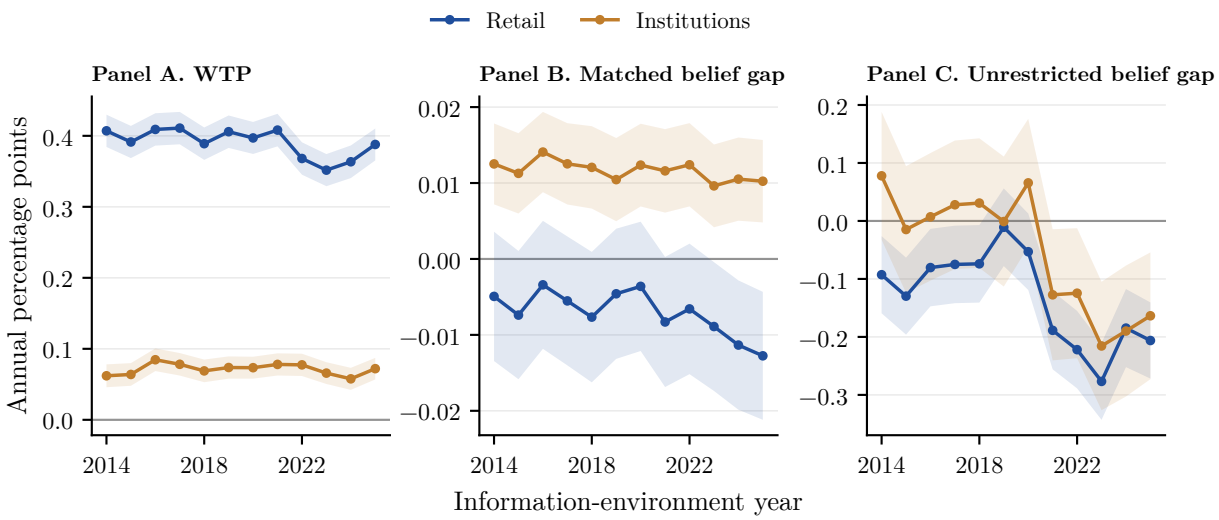


Figure B.8. Reconstructed series by investor class. This figure plots reconstructed willingness to pay (WTP) and belief gaps by investor class in ChatGPT-5.6 Sol across the 2014–2025 information environments, in annual percentage points. Panels A, B, and C report WTP, the matched belief gap, and the unrestricted belief gap, respectively. WTP averages the annual return given up for Portfolio H and the extra return required for Portfolio L; belief gaps are the expected return on H minus that on L. Both refer to annual returns over ten years. H and L score at the 80th and 20th ESG percentiles and are otherwise identical in the preference and matched-belief questions; unrestricted beliefs refer to typical real-world portfolios. The retail line averages the five political identities. The institutional line combines the large U.S. mutual fund company and the large U.S. pension fund. Their weights are the annual corporate-equity holdings of the mutual fund and ETF sector and of the pension sector in the Federal Reserve's Z.1 accounts (Board of Governors, 2026), normalized to sum to one. The mutual fund weight ranges from 0.68 to 0.74. Shaded bands are 95% confidence intervals, calculated as 1.96 reconstruction standard errors. For the combined lines, these standard errors account for the estimated correlation of errors across identities.

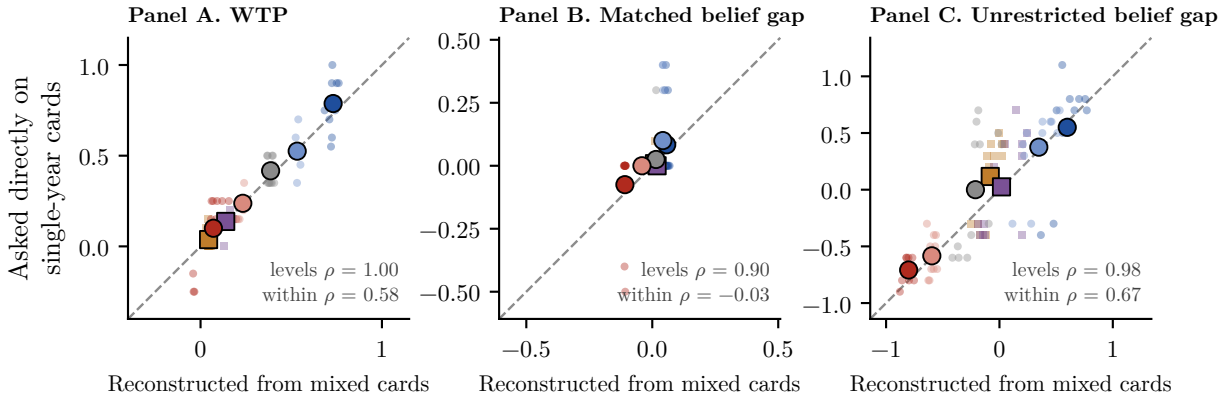


Figure B.9. Reconstructed versus direct single-year answers. This figure compares reconstructed responses with direct responses to the 12 single-year cards in ChatGPT-5.6 Sol, in annual percentage points. The horizontal axis shows reconstructed responses and the vertical axis direct responses. Panels A, B, and C report willingness to pay (WTP), the matched belief gap, and the unrestricted belief gap, respectively. WTP averages the annual return given up for Portfolio H and the extra return required for Portfolio L; belief gaps are the expected return on H minus that on L. Both refer to annual returns over ten years. H and L score at the 80th and 20th ESG percentiles and are otherwise identical in the preference and matched-belief questions; unrestricted beliefs refer to typical real-world portfolios. Large markers are identity means across the 12 years, and small markers are the 84 identity-year pairs. Circles denote the five political identities, using the colors of Figure B.7; squares denote the two institutions. The dashed line marks exact agreement. Each panel reports the correlation across identity means and the correlation within identities across years.

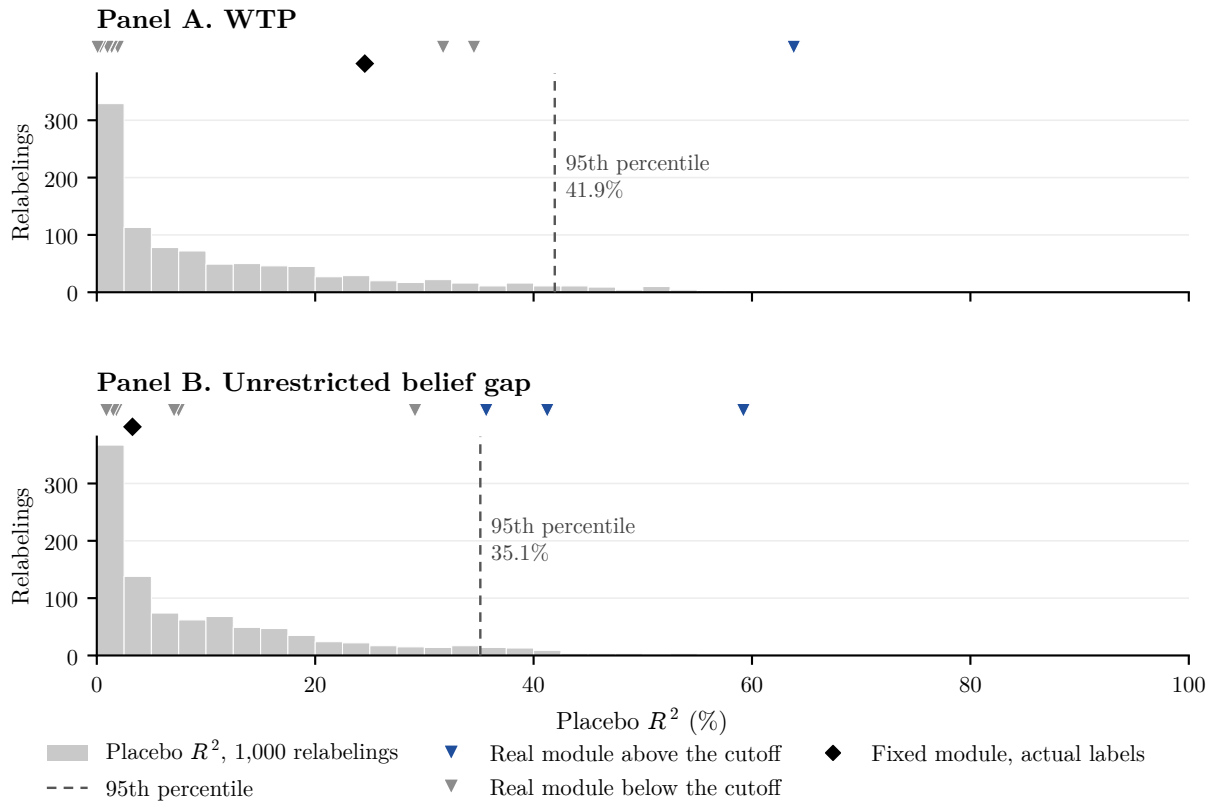


Figure B.10. Placebo distribution of the module R^2. This figure plots the placebo distributions of the module R^2 that set the cutoffs in Figure 11, in ChatGPT-5.6 Sol. In each of 1,000 draws, the year labels of the fixed tenth module are randomly reassigned across the 288 cards, with the same reassignment for all seven identities, and the ten-module version of equation (3) is re-estimated. The placebo R^2 is the R^2 of a univariate regression of that draw's aggregate reconstructed series on the relabeled module's annual effects over 2014–2025. The aggregate weights the seven identities by the fixed 2014 corporate-equity holdings of their sectors, as in Figure 11. Panel A shows willingness to pay (WTP), and Panel B shows the unrestricted belief gap. Bars are histograms of the 1,000 placebo values in bins of 2.5 percentage points. The dashed line marks the 95th percentile, the cutoff in Figure 11: 41.9% for WTP and 35.1% for the belief gap. Triangles above each histogram mark the R^2 of the nine real modules, blue if above the cutoff and gray otherwise. The black diamond marks the fixed module at its actual year labels.