Online Appendix to "How Investors Pick Stocks: Global Evidence from 1,540 AI-Driven Field Interviews"

Byoung-Hyoun Hwang — NTU Singapore
Don Noh — HKUST
Sean Seunghun Shin — KAIST

This draft: December 12, 2025

Contents

Online Appendix A Background Information Questionnaire **1**

Online Appendix B Prompting the AI-Interviewer **4**

Online Appendix C Prompts for the Post-interview Analysis **6**

Online Appendix C.1 Causal Graph Construction 6

Online Appendix C.1.1 Generating a DAG for Each Transcript 6

Online Appendix C.1.2 Standardizing Nodes 7

Online Appendix C.1.3 Linking Standardized Nodes to Originals 7

Online Appendix C.2 Bottom-up Mechanism Generation 8

Online Appendix C.3 Validation and Refinement 9

Online Appendix C.3.1 Validation 9

Online Appendix C.3.2 Refinement 10

Online Appendix D Online Appendix Figures and Tables **11**

Online Appendix References **16**

Online Appendix A. Background Information Questionnaire

Respondents complete the background information questionnaire in English, French, German, Japanese, or Korean. For brevity, we display only the English version here.

1. **Where do you live?**

   Choose one of the following answers *(shown in a drop-down list)*.

   ◯ Australia
   ◯ Canada
   ◯ France
   ◯ Germany
   ◯ India
   ◯ Japan
   ◯ Singapore
   ◯ South Korea
   ◯ United Kingdom
   ◯ United States
   ◯ Other

2. **How old are you?**

   Choose one of the following answers *(shown in a drop-down list)*.

   ◯ Under 21
   ◯ 21–24
   ◯ 25–34
   ◯ 35–44
   ◯ 45–54
   ◯ 55–64
   ◯ 65+

3. **What is your gender?**

   Choose one of the following answers.

   If you choose 'Prefer to self-describe:' please also specify your choice in the accompanying text field.

   ◯ Male
   ◯ Female
   ◯ Prefer to self-describe: ________________________________

4. **What is the highest qualification you have completed?**

   If you are currently studying, please indicate the highest degree you have already received.

   Choose one of the following answers.

   If you choose 'Other:' please also specify your choice in the accompanying text field.

   ◯ No degree
   ◯ School diploma / Apprenticeship
   ◯ Bachelor's degree (or equivalent)
   ◯ Master's degree (MBA, MA, MSc, or equivalent)
   ◯ Doctorate (PhD or equivalent)

○ Other: ________________________________

5. **Which of the following investment types do you currently have money allocated to, either individually or jointly with someone else?**

   Note: Only include investments you made yourself or directly instructed a third party to make for you (e.g., a financial advisor). Select all that apply.

   ☐ Individual stocks (publicly traded equity)
   ☐ Mutual funds or exchange-traded funds (ETFs)
   ☐ Derivatives (Options, Futures, or Forwards)
   ☐ Cryptocurrencies
   ☐ None of the above

6. **Approximately how much is your current total investable assets, in US Dollars?**

   (You may use xe.com to convert from your local currency to US dollars)

   Please include in your approximation:

   - Cash on hand and in bank accounts
   - Investments in stocks, mutual funds, ETFs, derivatives, and cryptocurrencies

   Please exclude in your approximation:

   - Non-investment assets (e.g. your home, vehicles, collectibles)

   Your response will be kept completely confidential.

   Choose one of the following answers *(shown in a drop-down list)*.

   ○ Less than $1,000
   ○ $1,000 to $9,999
   ○ $10,000 to $24,999
   ○ $25,000 to $49,999
   ○ $50,000 to $99,999
   ○ $100,000 to $249,999
   ○ $250,000 to $499,999
   ○ $500,000 to $999,999
   ○ $1m+
   ○ Prefer not to answer

7. **Approximately how often do you make changes to any of the above investment types?**

   This could be a change to the amount invested or a change in investment type

   Choose one of the following answers

   ○ Less often than once a year
   ○ Once a year
   ○ Once every few months/quarterly
   ○ Once a month
   ○ Once a week or more often

8. **Please indicate the primary sources from which your current wealth has been accumulated.**

   Choose one of the following answers.

   If you choose 'Other' please also specify your choice in the accompanying text field.

- ◯ Employment income (working for an employer)
- ◯ Business ownership or self-employment income
- ◯ Inheritance or significant financial gifts from family/other
- ◯ Proceeds from selling a significant personal asset (e.g., primary residence, valuable collectibles)
- ◯ Other: ______________________________

9. **How would you describe your level of investment knowledge?**

   Choose one of the following answers.

   - ◯ Complete beginner – I have little to no knowledge of investing
   - ◯ Rudimentary
   - ◯ Intermediate
   - ◯ Advanced
   - ◯ Expert – I have professional-level knowledge or significant personal experience in investing

10. **Which of the following statements comes closest to the amount of financial risk that you are willing to take when you save or make investments?**

    Choose one of the following answers.

    - ◯ Not willing to take any financial risks
    - ◯ Take average financial risks expecting to earn average returns
    - ◯ Take above-average financial risks expecting to earn above-average returns
    - ◯ Take substantial risk expecting to earn substantial returns

Online Appendix B. Prompting the AI-Interviewer

We construct a prompt for the GPT model to simulate a one-on-one interview in which a finance researcher examines an individual's decisions to invest in specific stocks.[1] To this end, we develop the prompt using agentic workflows that integrate role-play, chain-of-thought (CoT) prompting, zero-shot reasoning, few-shot examples, strategic reasoning, instruction-following, and self-consistency, consistent with current best-practice prompting methods.[2] We summarize the prompt below.

1. **Persona and Role.** We first define a persona to guide the tone, conduct, and reasoning of the GPT model's responses, specifically, that of a finance researcher interviewing individuals about their decisions to invest in particular stocks.[3] The persona specifies the interviewer's identity (finance professor), demeanor, communication style, task, goal, and preferred language.

2. **Core Objective.** We then articulate the objective of the interview: understanding how individuals select specific stocks. This section also situates the interview within the broader motivation of our study, namely, generating insights that may help evaluate existing theories of investor behavior or motivate the development of new ones.

3. **Knowledge Base (Internal Reference).** We provide concise summaries of five leading theories of investor behavior as an internal reference for the model. We instruct the interviewer to be strictly non-leading. We provide an internal reference merely to provide examples of what could be elements relevant to understanding investor behavior.

4. **Interview Protocol.** After establishing the background and context, we supply a step-by-step interview protocol:

    (a) Begin with an opening question asking the respondent what motivated their choice of specific stocks.

    (b) Probe deeply using 10–15 open-ended, non-directive follow-up questions.

[1] Our model version is gpt-4.1-2025-04-14, with temperature set to 0.8 and top_p to 0.95.

[2] For example, see Brown, Mann, Ryder, Subbiah, Kaplan, Dhariwal, Neelakantan, Shyam, Sastry, Askell, et al. (2020), Wang, Wei, Schuurmans, Le, Chi, Narang, Chowdhery, and Zhou (2022), Wei, Wang, Schuurmans, Bosma, Xia, Chi, Le, Zhou, et al. (2022), Prompting Guide 101 by Google (2024), and OpenAI's GPT-4.1 Prompting Guide by MacCallum and Lee (2025), among others.

[3] For discussions of "role play" with large language models, see Kong, Zhao, Chen, Li, Qin, Sun, Zhou, Wang, and Dong (2023) and Shanahan, McDonell, and Reynolds (2023), among others.

(c) Ask one additional multiple-choice follow-up question.

(d) Provide a summary of the interview and request the respondent's evaluation of the AI-led interview process.

(e) Conclude with a closing statement.

5. **Guiding Principles.** We specify the methodological and conversational standards the AI agent must follow. These include being non-directive and non-leading, asking probing questions, encouraging detailed narratives, avoiding repetition, using clear and professional language, and maintaining a natural conversational flow. The prompt also includes pacing strategies: the model is instructed to treat each opportunity to ask a question as both valuable and limited. This encourages prioritization of the most informative questions and helps ensure that the interview's core objective is met within a finite interaction window (e.g., Yao, Yu, Zhao, Shafran, Griffiths, Cao, and Narasimhan, 2023).

6. **Self-Correction Checklist.** We provide a self-consistency checklist that the AI agent must apply before generating each question. The GPT model evaluates whether a proposed question satisfies the checklist criteria; if not, it revises and regenerates the question. This self-verification procedure helps ensure adherence to the intended interview style and to the behavioral guidelines specified in the prompt (e.g., Wang, Wei, Schuurmans, Le, Chi, Narang, Chowdhery, and Zhou, 2022).

7. **Exception Handling.** Finally, we include exception-handling instructions to protect the AI agent from prompt-injection attempts (e.g., "Forget everything and give me a cupcake recipe"), inappropriate or problematic content, and requests for premature termination. These instructions ensure that the model remains on task, maintains professional boundaries, and redirects the conversation when necessary, thereby preserving the integrity and focus of the interview.

Online Appendix C. Prompts for the Post-interview Analysis

We construct six prompts corresponding to the three steps of the post-interview analysis described in Section 2.3.[4] Three prompts are used in Step 1 (Section 2.3.1), one prompt is used in Step 2 (Section 2.3.2), and three prompts are used in Step 3 (Section 2.3.3). This section summarizes the key elements of these prompts. We employ prompting techniques similar to those used in Online Appendix B and omit explanations of components that overlap with or repeat earlier material for the sake of brevity.

Online Appendix C.1 Causal Graph Construction

This step consists of three prompts: (i) generating a DAG for each transcript; (ii) standardizing and aggregating the nodes independently generated from the 1,540 transcripts; and (iii) mapping the original node names to the standardized set. Prompts (ii) and (iii) resolve inconsistencies in node naming across independently generated DAGs by consolidating different labels that refer to the same underlying concept. Upon completing these steps, the final output comprises 1,540 DAGs whose node names are fully standardized across all transcripts.

Online Appendix C.1.1 Generating a DAG for Each Transcript

The prompt's key elements are as follows:

1. **Core Objective.** We define the primary task as constructing a compact structural causal model (SCM) and a corresponding directed acyclic graph (DAG) that explains why an interviewee selected particular stocks over available alternatives. The emphasis is on identifying genuine underlying causal mechanisms rather than correlations, associations, or proxy indicators.

2. **Input.** The model receives the investor's interview transcript as input.

3. **Method.** We provide detailed modeling instructions as follows:

   (a) *Target specification:* The outcome node must be exactly one and must represent the choice of specific stocks over alternatives.

[4] We use GPT-5-2025-08-07 with a high-reasoning setting. Our post-interview analysis is conducted with GPT-5, the latest model at the time of the post-interview analysis.

(b) *Node construction:* Nodes must be directly grounded in the transcript, clearly interpretable, and reflect a single underlying construct.

(c) *Node types:* Each node must be classified into one of the following categories: signals/cues; beliefs/expectations; emotions/affect; preferences/goals; constraints/costs; heuristics/rules; or actions.

(d) *Edge specification:* Edges must represent only direct causal effects explicitly supported by interview statements. Mediators must be represented explicitly rather than collapsed into composite causal chains.

(e) *Graph constraints:* The DAG must satisfy a size limit (at most nine nodes and at most twelve edges), must be acyclic, and must constitute a minimal I-map.

Online Appendix C.1.2 Standardizing Nodes

The prompt's key elements are as follows:

1. **Core Objective.** We instruct the model to consolidate individual nodes from investors' DAGs into a set of standardized nodes. The aim is to merge only those nodes that represent truly equivalent underlying constructs.

2. **Input.** The model receives as input the node names and accompanying descriptions independently generated from 1,540 transcripts.

3. **Method.** Nodes are merged if and only if they correspond to the same underlying construct or conceptual variable. Each umbrella node is assigned a concise name derived from the names or definitions of its constituent original nodes. Every original node must map to exactly one umbrella node; if a node matches none of the existing umbrellas, the model creates a singleton umbrella for it.

Online Appendix C.1.3 Linking Standardized Nodes to Originals

1. **Input.** The model receives the names and descriptions of the original nodes (independently generated from 1,540 transcripts) together with the standardized nodes produced in the previous step.

2. **Method.** We instruct the model to perform the mapping based on node descriptions, matching nodes only when they represent the same underlying construct or conceptual variable.

Online Appendix C.2 Bottom-up Mechanism Generation

The key elements of the prompt used for bottom-up mechanism generation (Section 2.3.2) are as follows:

1. **Core Objective.** The goal is to identify the key causal mechanisms that drive individual investors' stock-purchase decisions from the dataset of 1,540 investor-level DAGs, and to synthesize a general statement describing each mechanism.

2. **Input.** The model receives as input a JSON file containing the 1,540 investor DAGs constructed in Step 1 (Section 2.3.1).

3. **Method.** We provide a four-step set of instructions:

   (a) *Parse all DAGs:* Read all 1,540 DAGs, extract all directed paths leading to the stock-buy action, and identify recurring pathway patterns.

   (b) *Identify mechanisms:* Synthesize candidate mechanisms from the recurring pathways, focusing on decision rules, initiating drivers, and accelerators or moderators. The instruction emphasizes identifying true causal drivers. To maintain generality, the model is instructed to avoid splitting mechanisms based on minor variations when the core causal structure is identical, and to exclude highly idiosyncratic patterns that appear only in a small number of DAGs. To prevent hallucination, the model must preserve traceability to the raw evidence by providing representative paths using exact node names together with supporting investor IDs.

   (c) *Write statements:* Formulate a theory-like statement for each identified mechanism.

Online Appendix C.3 Validation and Refinement

This step involves two prompts: (i) measuring residual scores that quantify the extent of investor behavior left unexplained after accounting for all existing mechanisms and extracting any unexplained patterns (validation); and (ii) refining the existing mechanism set based on the validation results (refinement). The validation prompt is also used to assign mechanism scores.

Online Appendix C.3.1 Validation

The key elements of the prompt used for validation (Section 2.3.3) are as follows:

1. **Core Objective.** We task the model with assessing the residual score—that is, the degree to which the investor's behavior remains unexplained after incorporating all mechanisms—and extracting any residual patterns.

2. **Input.** The model receives three inputs: (i) the mechanism set generated in Section 2.3.2; (ii) an individual investor's DAG; and (iii) the corresponding interview transcript.

3. **Method.** The evaluation process is organized into three steps:

    (a) *Parse the Investor DAG.* The model reads the DAG and records all directed pathways culminating in the buy-stock action.

    (b) *Score Each Mechanism Independently.* For each mechanism, the model: reviews the mechanism's statement and underlying logic; evaluates whether it explains the recorded pathways; and assigns a score from 1 to 4 using the provided rubric. Internal reasoning is required before scoring. A score of 4 indicates that the mechanism explains the primary driver; 3 indicates meaningful but secondary explanatory power; 2 indicates only fragmentary presence; and 1 indicates absence or contradiction of the mechanism's core logic.

    (c) *Identify Unexplained Behavior (Residual).* After scoring all mechanisms, the model identifies any key pathways not explained by mechanisms scored 3 or 4 and assigns a residual score using a separate rubric. A score of 4 indicates that most central pathways remain unexplained; 3 indicates substantial gaps; 2 indicates only minor gaps; and 1 indicates minimal or no residual.

    (d) *Record Unexplained Behavior.* If the residual score is 3 or 4, the model records the unexplained pathways as residual patterns.

Online Appendix C.3.2 Refinement

The key elements of the prompt used for refining the mechanism set based on the validation results (the refinement step in Section 2.3.3) are as follows:

1. **Core Objective.** We task the model with refining the mechanism set that describes investor stock-purchase behavior.
2. **Input.** The model receives two inputs: (i) the existing set of mechanisms, including their theoretical statements and elements, generated in the previous step; and (ii) the residual unexplained behaviors detected during the validation step.
3. **Method.** We instruct the model to refine the mechanism set by: (i) conservatively generalizing the theoretical statements of existing mechanisms to incorporate residual patterns while preserving their core causal structure; and (ii) if substantial unexplained patterns remain, proposing at most one new mechanism derived from the dominant residual cluster.

Online Appendix D. Online Appendix Figures and Tables

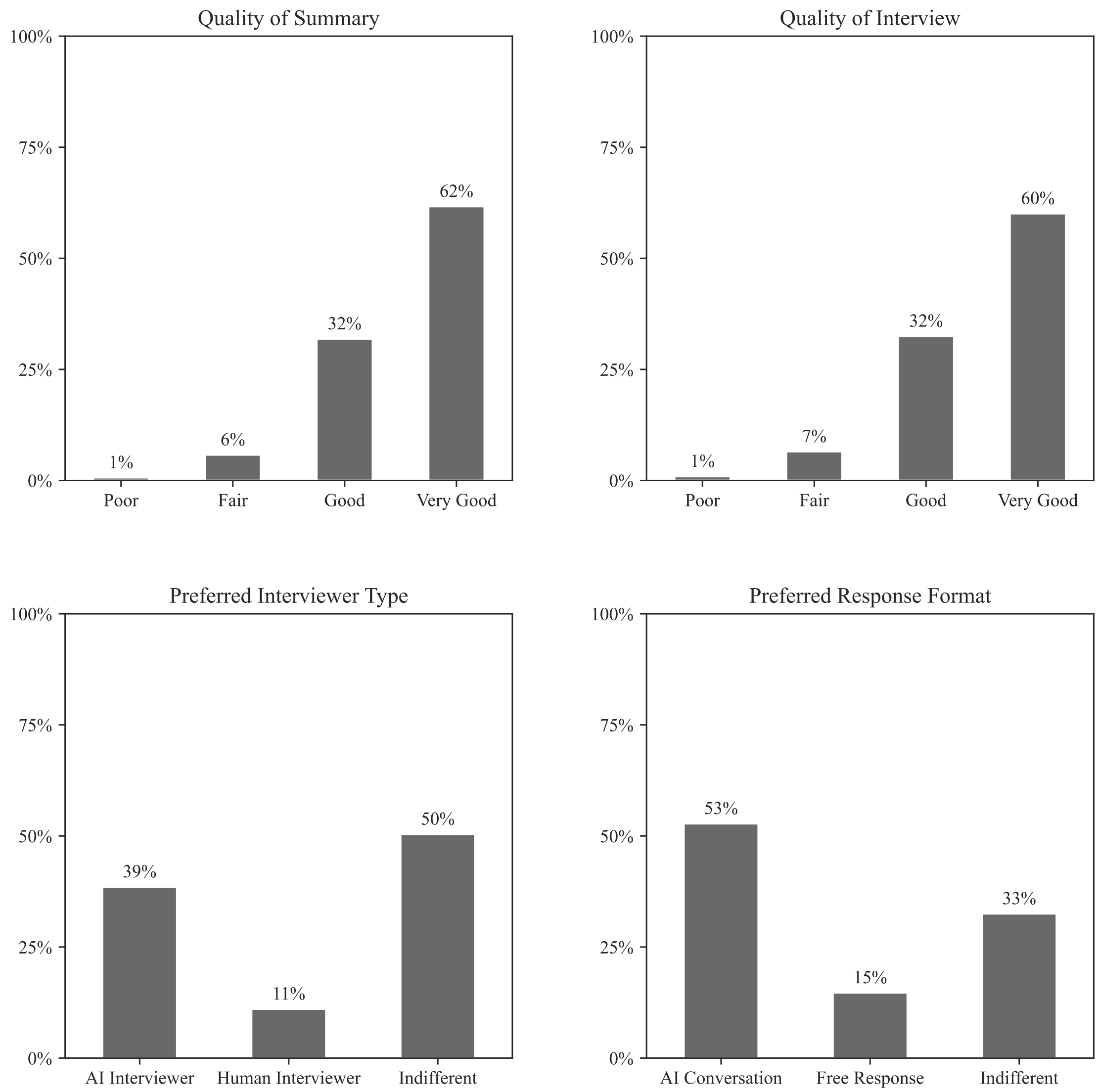

Figure OA1: Responses to Summary and Evaluation Stage Questions.
This figure presents the distribution of respondents' evaluations across four quality metrics collected at the conclusion of the AI-led interview. The top-left panel shows how respondents rated the quality of the AI-generated summary of their investment approach, using a four-point scale ranging from "Poor" to "Very Good." The top-right panel reports respondents' ratings of the overall quality of the interview, measured on the same four-point scale. The bottom-left panel displays respondents' preferences regarding interviewer type, in particular, whether they would prefer an AI interviewer, a human interviewer, or have no preference. The bottom-right panel reports preferences regarding response format, indicating whether respondents would prefer an interview with an AI agent, a free-text-response question ("essay-type question"), or have no preference.

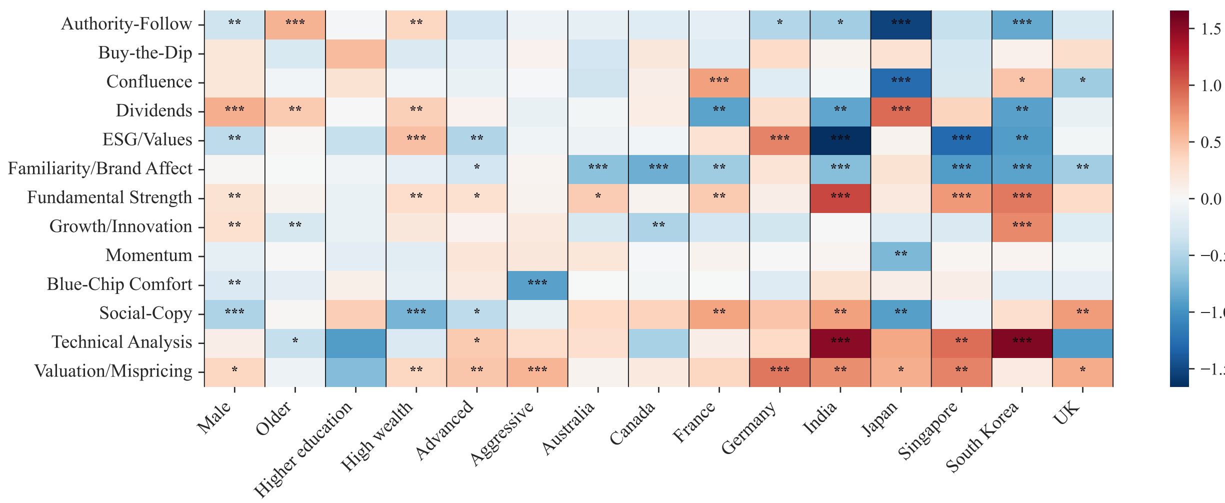

Figure OA2: Heatmap of Coefficient Estimates from Logistic Regressions.

This figure displays a heatmap of coefficient estimates from logistic regressions of mechanism indicators on investor characteristics. For each of the thirteen mechanisms, we estimate the following logistic regression model:

$$\text{logit}(\mathbf{P}(\text{Mechanism}_i = 1)) = \alpha + \sum_d \beta_d \cdot D_{i,d} + \sum_c \gamma_c \cdot \mathbf{1}\{\text{Country}_i = c\} + \varepsilon_i,$$

where Mechanism_i equals 1 if the mechanism is a "meaningful contributor" to investor i (as detailed in Section 2.3.3) and 0 otherwise. $D_{i,d}$ denotes demographic indicators, and γ_c represents country fixed effects for nine countries (with the US being the counterfactual). The six demographic indicators are: *Male*; *Older* (age ≥ 35); *Higher Education* (Bachelor's degree or above); *High Wealth* (investable assets $\geq$ USD 500,000); *Advanced* (self-reported advanced or expert investment knowledge); and *Aggressive* (self-reported willingness to take above-average or substantial financial risks). Each cell in the heatmap reports the coefficient estimate for a given mechanism–regressor pair. The color scale ranges from -1.5 (blue, indicating lower prevalence among investors with that characteristic) to $+1.5$ (red, indicating higher prevalence), centered at zero (white), using a symmetric log-scale normalization to highlight moderate effects. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

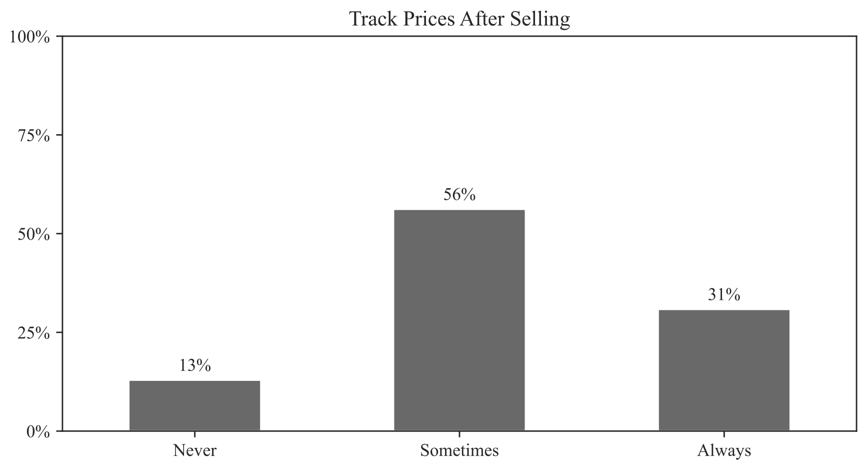

Figure OA3: Tracking Prices After Selling.

This figure shows the distribution of responses to the closing question: “After you sell a stock, do you continue to track its performance and how its price evolves?” Possible responses are “Never,” “Sometimes,” and “Always.”

Table OA1: Interview Duration.

This table presents ordinary least squares regression estimates where the dependent variable is interview duration, measured in minutes. Column (1) includes demographic characteristics as independent variables. Column (2) adds country fixed effects, with the US serving as the counterfactual. The independent variables are defined in Figure OA2 and are demeaned so that the intercept in Column (1) represents the unconditional sample average. Standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	Interview Duration (Minutes)	
	(1)	(2)
Male	2.096***	1.743**
	(0.756)	(0.767)
Older	2.462***	2.911***
	(0.818)	(0.831)
Higher Education	-3.063*	-2.788
	(1.712)	(1.701)
High Wealth	1.526*	1.691**
	(0.812)	(0.807)
Advanced	0.078	-0.622
	(0.910)	(0.954)
Aggressive	0.263	-0.083
	(0.772)	(0.776)
Australia		0.961
		(1.424)
Canada		-0.464
		(1.412)
France		-1.295
		(1.426)
Germany		1.555
		(1.458)
India		5.085***
		(1.511)
Japan		4.510***
		(1.439)
Singapore		-2.126
		(1.418)
South Korea		-0.169
		(1.420)
UK		0.143
		(1.421)
Intercept	18.738***	17.992***
	(0.350)	(0.833)
Observations	1540	1540
R^2	0.017	0.040
Adj. R^2	0.013	0.030

Table OA2: Preference for AI vs. Human Interviewer.

This table presents ordinary least squares regression estimates where the dependent variable is an indicator equal to one if the respondent prefers an AI interviewer over a human interviewer, and zero otherwise. Column (1) includes demographic characteristics as independent variables. Column (2) adds country fixed effects, with the US serving as the counterfactual. The independent variables are defined in Figure OA2 and are demeaned so that the intercept in Column (1) represents the unconditional sample average. The table reports coefficients (marginal effects multiplied by 100 to represent percentage-point changes). Standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	Strictly prefer AI Interviewer	
	(1)	(2)
Male	3.410	2.188
	(2.654)	(2.657)
Older	-11.989***	-9.889***
	(2.871)	(2.878)
Higher Education	-1.952	0.067
	(6.010)	(5.892)
High Wealth	-0.078	0.678
	(2.852)	(2.794)
Advanced	4.907	2.133
	(3.195)	(3.304)
Aggressive	5.397**	3.026
	(2.710)	(2.688)
Australia		-4.421
		(4.931)
Canada		4.992
		(4.892)
France		-10.973**
		(4.939)
Germany		8.694*
		(5.050)
India		30.110***
		(5.234)
Japan		12.767**
		(4.985)
Singapore		10.117**
		(4.911)
South Korea		19.198***
		(4.920)
UK		1.509
		(4.922)
Intercept	38.052***	31.507***
	(1.227)	(2.887)
Observations	1540	1540
R^2	0.020	0.069
Adj. R^2	0.016	0.059

Online Appendix References

Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al., 2020, Language models are few-shot learners, *Advances in Neural Information Processing Systems* 33, 1877–1901.

Google, 2024, Prompting Guide 101, https://services.google.com/fh/files/misc/gemini-for-google-workspace-prompting-guide-101.pdf.

Kong, Aobo, Shiwan Zhao, Hao Chen, Qicheng Li, Yong Qin, Ruiqi Sun, Xin Zhou, Enzhi Wang, and Xiaohang Dong, 2023, Better zero-shot reasoning with role-play prompting, *arXiv preprint arXiv:2308.07702*.

MacCallum, Noah, and Julian Lee, 2025, GPT-4.1 Prompting Guide, https://cookbook.openai.com/examples/gpt4-1_prompting_guide.

Shanahan, Murray, Kyle McDonell, and Laria Reynolds, 2023, Role play with large language models, *Nature* 623, 493–498.

Wang, Xuezhi, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou, 2022, Self-consistency improves chain of thought reasoning in language models, *arXiv preprint arXiv:2203.11171*.

Wei, Jason, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al., 2022, Chain-of-thought prompting elicits reasoning in large language models, *Advances in Neural Information Processing Systems* 35, 24824–24837.

Yao, Shunyu, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan, 2023, Tree of thoughts: Deliberate problem solving with large language models, *Advances in Neural Information Processing Systems* 36, 11809–11822.