What Would Jeeves Do? Inflation Expectations, AI Advice, and Household Spending*

Yuriy Gorodnichenko[a] and Philipp Heil[b]

[a]UC Berkeley, NBER, and CEPR
[b]HEC Paris and ifo Institute

November 3, 2026

This is a preliminary draft. Please do not cite or circulate without the permission of the authors.

Abstract

Using a large survey, we provide new evidence on how Americans use generative AI for macroeconomic information and advice. In a randomized controlled trial, we generate exogenous variation in inflation expectations and in access to a conversation with an AI assistant, which lets us study how interactions with AI shape behavioral responses to expectations. Respondents mostly ask about their household budgets, and the AI generally recommends cutting back on expenditures. We show that this advice affects respondents' planned and actual spending. At the same time, the conversation leaves respondents' uncertainty, narratives about inflation, and understanding of economic concepts largely unchanged, pointing to a distinct friction: households may know what will happen without knowing what to do about it.

JEL Classification: C83, D14, D84, E21, E31, O33
Keywords: inflation expectations, generative AI, household spending, financial advice, randomized controlled trial

*We are grateful for feedback and comments from Joe Hazell and Ryan Oprea. We thank Leander Wolff for outstanding research assistance. This experiment was pre-registered at the AEA RCT Registry under the RCT ID: AEARCTR-0019094 and received IRB Approval by the University of California, Berkeley, Committee for the Protection of Human Subjects (CPHS) (protocol #2026-06-19748).

1 Introduction

In 1829, the German poet Johann Wolfgang von Goethe famously observed, "Knowing is not enough; we must apply. Willing is not enough; we must do." Two hundred years later, this maxim aptly summarizes a central tension in economics: expectations sit at the heart of modern models, and yet the transmission from beliefs to actions remains a puzzle. In standard models, decisions respond strongly to changes in, say, expected inflation or policy rates. In practice, the pass-through from beliefs to actions is present but considerably weaker.[1]

Why is the pass-through relatively weak? Our hypothesis is that households may not know what to do with their beliefs. For instance, knowing that inflation will be higher is not the same as knowing whether this calls for buying a car now, paying down a mortgage, or shifting savings out of a bank account. Faced with a large menu of possible actions and little sense of which one fits their circumstances, households may end up acting on none of them. In this case, guidance might be the natural remedy for inaction: if changes in expectations are coupled with guidance on the course of action (or at least with guidance that narrows the set of options), the transmission from beliefs to actions should change, in a direction that depends on the content of the guidance.

To test this hypothesis, we run a randomized controlled trial (RCT) to generate independent exogenous variation in households' beliefs and in the advice they receive about what those beliefs imply. In a survey of approximately 4,000 Americans recruited through Prolific in July 2026, we randomly assign respondents to one of four arms. One group of respondents is told about actual past inflation. Another group holds a conversation of up to five minutes with an AI assistant about inflation and its implications for their spending and saving. A third group receives both the inflation information and the AI conversation. The control group receives neither. After treatment, we re-elicit beliefs and spending plans and ask respondents to allocate a hypothetical windfall across assets, to report how they would respond to a hypothetical inflation surprise, and to answer questions about their understanding of inflation. About one month later, we re-contact respondents to measure realized spending and the persistence of beliefs.

The novel element of our design is the AI conversation, and we built it to approximate the advice households can obtain in the "wild." Rather than following a script, the chatbot responds adaptively to whatever respondents ask. It is instructed only to discuss the implications

[1] Perhaps the most striking illustration of the gap between theory and evidence is forward guidance: under full-information rational expectations (FIRE), a credible promise about policy rates far in the future should generate huge effects on current spending, yet its measured effects are far smaller. More directly, Coibion et al. (2020) find that firms raise their prices by about 25 basis points in response to a one percentage point increase in their inflation expectations, well short of the near one-for-one response implied by theory. Giglio et al. (2021) report that the sensitivity of investors' portfolios to their return expectations is a fraction of what standard models predict.

of inflation for the respondent's spending and saving in general terms, and it is given the respondent's demographic information so that it can frame its answers around their own circumstances. We record every conversation and classify its content, which lets us document what households ask an AI assistant about inflation and what advice they receive. In addition to providing advice to households, the AI conversations thus give us a window into what information households seek about inflation and its implications.

We document several key findings. First, AI is already an important source of economic information for households: one in five respondents report using an AI assistant for economic topics on a weekly basis, and respondents consult AI assistants about as frequently as TV, radio, and print, and more frequently than government statistics or financial advisors. Households trust AI most for informational tasks, such as understanding a concept or getting a current figure, and least for high-stakes decisions such as choosing a mortgage or changing investments.

Second, when given the opportunity, households use the AI assistant to ask about their own finances rather than about the macroeconomy. Budgeting, saving, debt, and purchases are the primary topics: 80% of respondents ask about their own financial decisions, against 28% who ask about the inflation outlook and 19% who ask how inflation works. The advice households receive is practical and cautious. The AI recommends budgeting and prioritizing essentials (89% of respondents) and cutting back (61%), while advice to spend more is rare apart from buying ahead of price increases (22%). The conversations look the same whether or not respondents saw the official inflation number beforehand. Respondents rate the conversations as helpful, and 80% report having learned something. At the same time, having a conversation with the AI does not materially change average inflation forecasts and lowers inflation uncertainty only slightly.

Third, and most important, the AI conversation changes how beliefs translate into decisions. Using the treatments as instruments for posterior beliefs, we find that a 1 percentage point increase in expected inflation raises planned spending on durable goods by about 0.19 log points *on average*, consistent with intertemporal substitution, and lowers the planned share of risky assets in a hypothetical portfolio. The durable-spending pass-through is concentrated among households who did not talk to the AI. For respondents without the AI conversation, the coefficient is 0.21. For respondents who had the conversation, it is -0.02. Because 2SLS scales each spending response by how much beliefs moved, this difference cannot be attributed to differences in how strongly the arms shift expectations. Instead, it suggests that the advice households received, which largely counseled belt-tightening, offset the response that their revised beliefs would otherwise have produced. Although estimates of pass-through into realized spending are less precise, we observe that the direction of AI advice can affect the direction of spending.

Fourth, the leading alternative explanations for weak pass-through do not account for this pattern. The treatments change cognitive uncertainty about decisions rather modestly. They have little effect on respondents' narratives about the causes of inflation (apart from a modest decline in the share blaming general government policy). They do not improve comprehension of inflation or financial literacy. While respondents report a median inaction band of 4 percentage points (that is, the increase in inflation they say would be needed before they change their behavior), the treatments do not shift this threshold, so thresholds cannot explain why the arms act differently. These results suggest that, conditional on expectations, what households do depends on what they are told to do.

These results have direct implications for policy. Central bank communication with the public is almost entirely about the current inflation rate, the inflation target, and the expected path of policy rates (Blinder et al., 2024). Our findings suggest that this may not be enough. Moving households' expectations could be relatively easy, but the decisions that follow depend on how households interpret what the new expectations mean for them. If households need guidance on what to do, and not only on what is happening, then the content of that guidance matters for the strength of the expectations channel.

We also note that as more households turn to AI for economic advice, the design choices of AI providers, such as system prompts, safety guidelines, and default framings, may come to shape the aggregate response of spending to inflation news. The same technology can give policymakers a scalable way to explain not only the outlook but also its implications for ordinary decisions. Understanding how AI advice interacts with expectations is thus relevant both for the transmission of monetary policy and for the regulation of AI tools that offer financial advice.

Our paper contributes to several literatures. First, we provide, to our knowledge, the first systematic evidence on how households use AI as a source of macroeconomic information. Existing work documents the rapid adoption of generative AI at work and in daily life (e.g., Board of Governors 2026; Pew Research Center 2026) and its growing use for personal finance (e.g., FINRA Investor Education Foundation 2024; Gallup 2026; Financial Conduct Authority 2026). We show that AI assistants are already an established source of information about the macroeconomy: 57% of respondents consult them at least once a month, more often than official statistics or financial advisors. Households trust AI most for informational tasks and least for high-stakes decisions, and usage differs systematically across demographic and partisan groups.[2]

Second, our paper relates to the large body of work on how households form macroeconomic

[2]Economists have also begun to use AI as a research tool, to conduct interviews (Geiecke and Jaravel, forthcoming; Chopra and Haaland, 2023; Chopra et al., 2026) and to simulate survey respondents (Horton, 2023). Recent research also shows that LLMs are already fairly good predictors of household behavior (Huang et al., 2026). We instead study AI as a source of information and advice for households.

expectations and how they respond to information. Survey evidence documents pervasive departures from full-information rational expectations, or FIRE (Coibion and Gorodnichenko, 2012, 2015; Bordalo et al., 2020). Randomized information treatments show that households' inflation expectations respond strongly to publicly available statistics (see Haaland et al., 2023 and Coibion and Gorodnichenko, 2026 for reviews). We show that an interactive AI conversation about inflation has a much weaker effect on inflation expectations than a simple statistic does.

Third, we contribute to a growing literature that uses RCTs to estimate the causal effect of expectations on decisions (see Coibion and Gorodnichenko, 2026 for an overview). A recurring finding is that the pass-through from beliefs to actions is weaker than models predict, which echoes observational evidence on the weak sensitivity of decisions to beliefs (e.g., Bachmann et al., 2015). Our contribution is to create exogenous variation not only in beliefs but also in guidance on the implications of inflation. This allows us to show that the strength of pass-through is not a fixed structural parameter. It depends on the advice households receive, holding expectations fixed.

Fourth, our results speak to explanations for why the pass-through from expectations to actions can be weak. Proposed mechanisms include uncertainty (Brainard, 1967), subjective models and narratives of the economy (e.g., Andre et al., 2022, Candia et al., 2020, Stantcheva, 2024), cognitive uncertainty (Enke and Graeber, 2023), limited financial literacy and money illusion (e.g., Shafir et al., 1997, D'Acunto et al., 2023), and fixed costs of adjustment (e.g., Stokey, 2009, Andersen et al., 2020). Theoretical work has modeled departures from FIRE that dampen the effects of forward guidance (e.g., Angeletos and Lian, 2018, Gabaix, 2020). Because the AI conversation changes decisions without materially changing cognitive uncertainty, narratives, or comprehension, our evidence points to a further, complementary friction: households may not know what the right action is, even when they know what will happen.

Finally, we contribute to the literature on financial advice. Human advisors can improve (or distort) household decisions (e.g., Mullainathan et al., 2012), and robo-advisors can shift portfolio choices (e.g., D'Acunto et al., 2019). A rapidly growing literature now studies advice from generative AI, mirroring the fast adoption of these tools by households. Choukhmane et al. (2026) simulate the lifetime effects of following GPT-5.2's advice and find that it would move households toward life-cycle prescriptions, though recommendations vary with what users write. Others randomize the advice itself: AI recommendations pull fund and pension choices toward whatever they promote (Greubel et al., 2026; Choi et al., 2026), and a chatbot led to worse savings-account choices than a standard Q&A page without changing comprehension (Ahmed et al., 2025). Similar to this earlier work, we find that AI advice moves choices more than understanding, but our paper differs from this work in three ways.

First, rather than delivering fixed recommendations or instructing an assistant to promote an option, we randomize access to an open-ended conversation with a neutrally instructed general-purpose assistant, so the advice reflects the questions *households* choose to ask and what the assistant recommends unprompted. Second, we study macroeconomic news rather than product or portfolio choice. Third, rather than simulating the effects of following advice, we measure the effects of receiving it on beliefs, planned spending and portfolios, and spending a month later. To our knowledge, ours is the first RCT to trace how an open-ended conversation with an AI assistant shapes responses to macroeconomic news, from what households ask to what they believe, plan, and do.

The rest of the paper is organized as follows. Section 2 describes the survey and the experimental design. Section 3 documents households' use of AI for macroeconomic information, Section 4 their prior beliefs, and Section 5 their conversations with the AI assistant. Section 6 presents treatment effects on beliefs. Section 7 estimates the pass-through from expectations to decisions, shows how it depends on AI advice, and examines mechanisms. Section 8 concludes, and the appendix contains additional results, the survey instrument, and the assistant's system prompt.

2 Survey Description

2.1 Sample and implementation

We fielded the main survey on Prolific on July 7, 2026, targeting $N = 4{,}000$ U.S. residents. Respondents who did not consent or who failed an attention check were screened out and not paid. Of the 4,031 completed responses, we exclude in total 161 respondents who had a missing or duplicate Prolific ID, failed the attention check, were flagged by a typing-speed bot filter (response text arriving faster than 10 characters per second), completed the survey in under one-third of the median completion time, took implausibly long on individual questions or on the survey as a whole, or gave invalid answers to the hypothetical-scenario or density-forecast questions.[3] The final analysis sample consists of 3,870 respondents (demographic composition is described in Appendix Table A1). The median (mean) respondent completed the survey in 23.9 (25.3) minutes.

A growing concern with online panels is that respondents answer with the help of LLMs, or that automated agents fully answer surveys (Westwood, 2025). Our design addresses this concern in two ways. First, the typing-speed filter flags text entered faster than humans can reasonably type. Second, the chat interface disables pasting, so respondents in the chatbot arms cannot generate their messages with an external AI tool. The typing-speed filter flags only 0.7% of the raw sample, which is consistent with recent evidence that bot infiltration on

[3] The bot, speeding, and duration criteria were pre-registered prior to data collection.

Prolific remains low (Celebi et al., 2026).

Figure 1 summarizes the structure of the survey (see Appendix B for the survey instructions). Respondents first report their age, gender, education, household income, employment status, political affiliation, and monthly non-durable spending. We then ask how often, for which topics, and for what purposes respondents use AI tools to inform themselves about the economy, how accurate and trustworthy they find AI-generated economic information, and whether AI advice changed their view on a topic. Respondents also describe their sources of macroeconomic information. Respondents next report their prior inflation expectations for the next 12 months, the main drivers of current inflation, their durable purchase plans, their confidence in making optimal spending decisions, and their preferred source of advice on inflation together with their willingness to pay for it. Respondents are then randomly assigned to receive information on recent CPI inflation, a conversation with our AI assistant, both, or neither (Section 2.2).

After the treatment, we elicit posterior inflation expectations and uncertainty, and respondents describe in their own words the main driver of inflation, how the government should respond, and the consequences of inflation for their household. Respondents then report their planned spending on durable and non-durable goods, their certainty that these plans are optimal, and how they would allocate a hypothetical $10,000 windfall. We next ask how respondents would adjust their purchases and behavior if inflation turned out 2 percentage points higher or lower than they expected. The survey closes with questions on the comprehension of inflation-related concepts and on financial literacy.

2.2 Design of Treatment Interventions

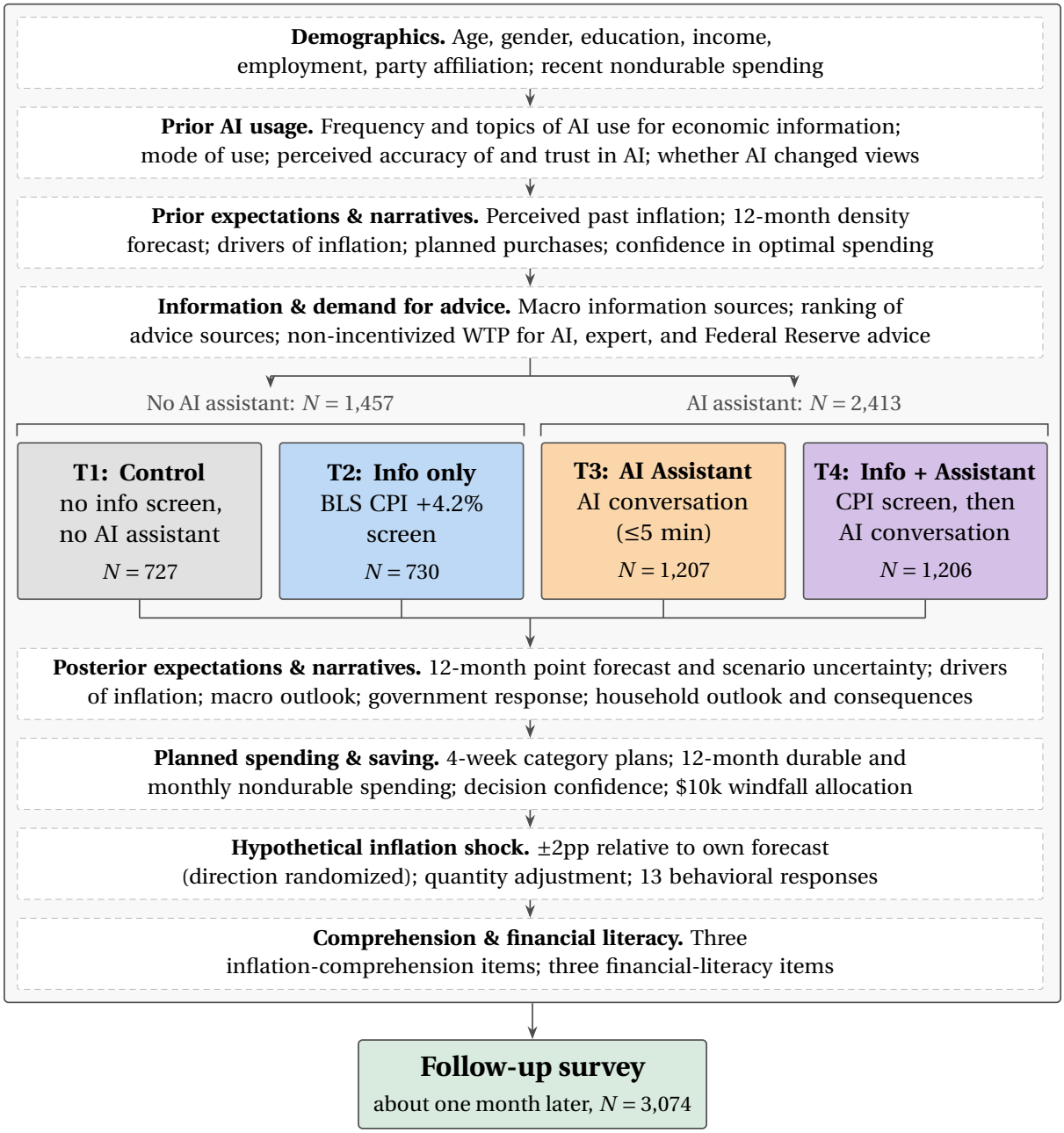
We embed two randomized interventions in the survey. The first is a standard information treatment (Coibion and Gorodnichenko, 2026), designed to generate exogenous variation in inflation expectations. Respondents in the information arms read the following statement:

> *Before you proceed, we would like you to know that, according to the U.S. Bureau of Labor Statistics, the Consumer Price Index increased by 4.2% over the last 12 months.*

The second intervention is a conversation of up to five minutes with an AI assistant that discusses the implications of inflation for respondents' spending and saving decisions. Respondents who receive both interventions see the information screen first. Section 2.3 describes the design and implementation of the AI assistant in more detail.

The two interventions are assigned independently in a 2×2 factorial design. After the exclusions described in Section 2.1, the four cells contain 727 respondents in the control group (T1), 730 in the information-only arm (T2), 1,207 in the AI-assistant-only arm (T3), and

Figure 1 Overview of the Experimental Design



1,206 in the arm receiving both (T4); see Figure 1. Assignment is balanced on pre-treatment covariates (Appendix Table A2). We oversample the AI assistant arms to increase power for the analysis of conversation transcripts, which are available only for these respondents. The information dimension is balanced, with 1,936 respondents receiving the information treatment and 1,934 not receiving it. Comparing each treated cell with the control group identifies the effect of the information treatment alone (T2), the AI assistant alone (T3), and their combination (T4). The combined cell helps us understand potential interaction effects of the two treatments.

2.3 The AI chatbot intervention

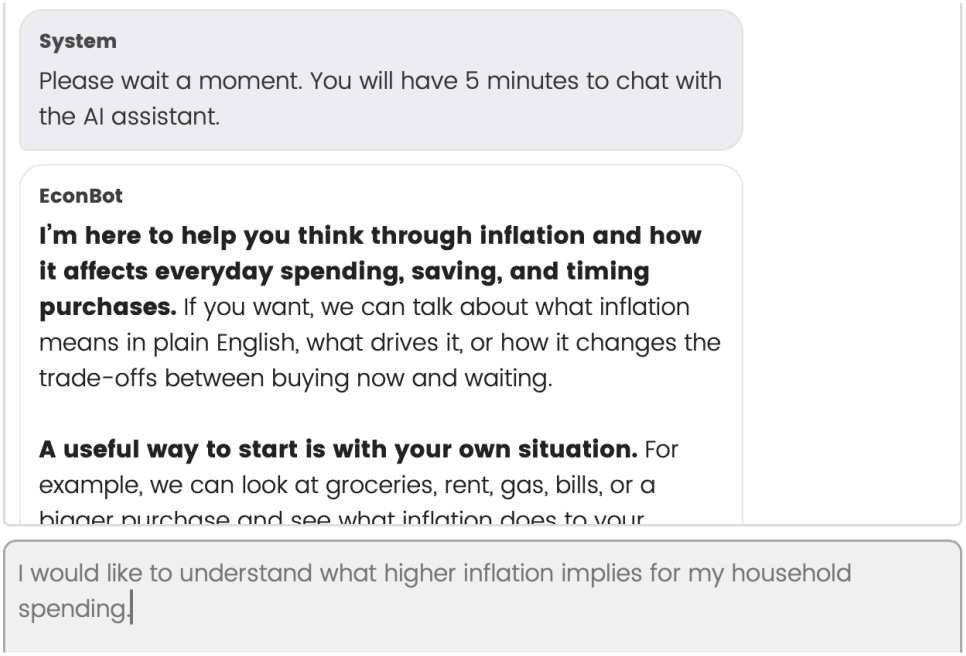
We design the AI assistant intervention to resemble how households interact with AI chatbots outside the survey, while keeping the conversation focused on the implications of inflation for spending decisions. To make the discussion specific to each respondent rather than to a generic household, the opening message of each conversation passes the respondent's age, gender, education, income, employment status, and non-durable spending, all elicited earlier in the survey, to the AI model. We take three measures to keep conversations on topic. First, respondents learn before the conversation that their messages will be scored for engagement and fit with the topic, and that the top 20% will receive an additional $1.[4] Second, the input box is pre-filled with the message "I would like to understand what higher inflation implies for my household spending," which respondents can edit before sending; ≈60% of respondents in the AI assistant arms send it unchanged. Third, the system prompt instructs the assistant to acknowledge off-topic messages in a brief clause and redirect the conversation to inflation and household spending, connecting the question to this topic where a natural bridge exists, for example from wages to real purchasing power (Appendix C).

Response Content and Style We customize the content and presentation of the assistant's replies through a tailored system prompt (Appendix C). The content is governed by three principles. First, the assistant matches each reply to the type of question rather than following a fixed template: it answers factual questions directly, illustrates conceptual questions with a short concrete example, and, for decision-related questions, states the trade-off explicitly and proposes one concrete next step, using the respondent's background information to make examples and trade-offs specific. Second, it grounds its guidance in mainstream economic reasoning, frames it neutrally rather than as a personal recommendation, presents the considerations on both sides of a decision, and gives no investment advice. Third, it stays within a defined scope: inflation, its causes and measurement, interest rates, purchasing power, and their implications for spending, saving, and durable purchases. These restrictions change the topics of the replies only modestly. When we compare a response generated under our system prompt with one generated by GPT-5.4 Mini without a custom prompt, both cover the two topics that dominate actual conversations, prices and everyday budgeting, but differ on less common topics (Appendix D).

The prompt also fixes the presentation, so that replies have a consistent structure across messages and respondents. Each reply consists of at most three short paragraphs of plain prose, with the key point of each paragraph in bold, followed where appropriate by a "Your next step" block and at most one follow-up question. This structure resembles the default presentation of general-purpose chatbots such as ChatGPT or Claude.

[4] GPT-5.6 Luna scores each respondent's messages on a 0–100 scale. Because the bonus rewards engagement and topical fit rather than agreement with any particular view, it should not favor any direction of belief revision.

Figure 2 Chatbot interface in main survey



Notes: Screenshot of the chat interface shown to respondents in T3 and T4. All chatbot-arm respondents face the same model, system prompt, and temperature.

Implementation We built our own AI assistant infrastructure and embedded it in a standard Qualtrics survey (Figure 2). Respondents could type in a text box or record voice messages, which were transcribed automatically. Each message from the respondent was sent to the OpenAI API, which generated a response with GPT-5.4 Mini at a low temperature (to reduce randomness in the responses), conditional on the system prompt.

2.4 Follow-up survey

About one month after the main survey, we recontacted all respondents who completed it for a shorter follow-up survey. The follow-up re-elicits inflation expectations and the response to the hypothetical inflation shock, and measures realized spending, comprehension, and the use of AI for household finance; respondents also complete a second conversation of up to three minutes with the AI assistant. These data let us assess whether the treatment effects on beliefs and decisions persist and whether the treatments change respondents' use of and attitudes toward AI. The follow-up was open from August 6 to August 31, 2026. Of the 3,870 respondents in the main-survey analysis sample, 3,074 (79.4 percent) completed the follow-up. Attrition does not differ across treatment arms.

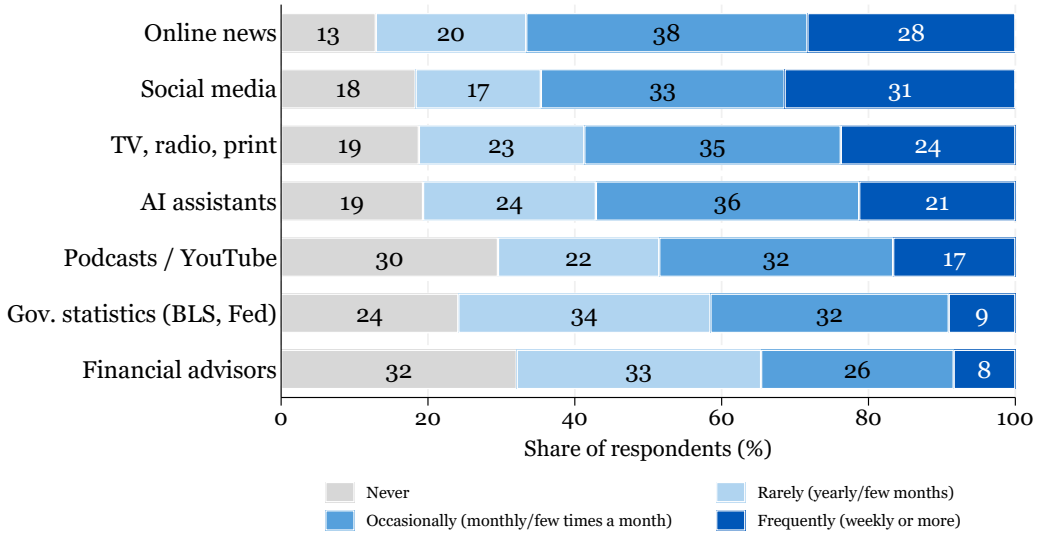
3 AI as a source of macroeconomic information

In this section, we provide the first systematic evidence of how households use AI as a source of macroeconomic information. To this end, we use responses to the pre-treatment questions

on AI usage. To complement existing evidence on the adoption of generative AI at work and in daily life (Humlum and Vestergaard, 2025; Chatterji et al., 2025; Bick et al., 2026; Board of Governors, 2026; Pew Research Center, 2026) and on its use for personal finance (FINRA Investor Education Foundation, 2024; Gallup, 2026; Financial Conduct Authority, 2026; Bilal et al., 2026), we focus on how households use AI to learn about the macroeconomy. Six facts emerge from this observational evidence.

Fact 1: AI assistants are already an established source of economic information, but their use is heterogeneous. Asked how often they consult various sources of macroeconomic information, 57% of respondents report using AI assistants at least once a month and 21% weekly or more, comparable to TV, radio, and print, below only online news and social media, and more often than official statistics or financial advisors (Figure 3). Use for economic topics more broadly is even more frequent: 40% of respondents use AI for economic topics a few times a week or more, while 39% use it rarely or never (Panel A, Figure 4), which suggests that much of this use concerns personal finances rather than the macroeconomy. These shares are high relative to earlier surveys of the general population, which measure AI use for any purpose (e.g., Bick et al. 2026; Pew Research Center 2026).

Figure 3 Macroeconomic Information Sources Used



Notes: Self-reported frequency of consulting each source of macroeconomic information, elicited before treatment. $N = 3{,}870$.

Fact 2: Respondents view AI as at least as accurate as most alternative sources (Panel B, Figure 4). Among respondents with an opinion, 60% consider AI more accurate than family and friends and 43% consider it more accurate than news media, against 18% and 19% who consider it less accurate. Relative to the government and central bank, the shares are balanced (31% each), and only financial professionals are rated more accurate than AI by more respondents than the reverse (42% against 28%).

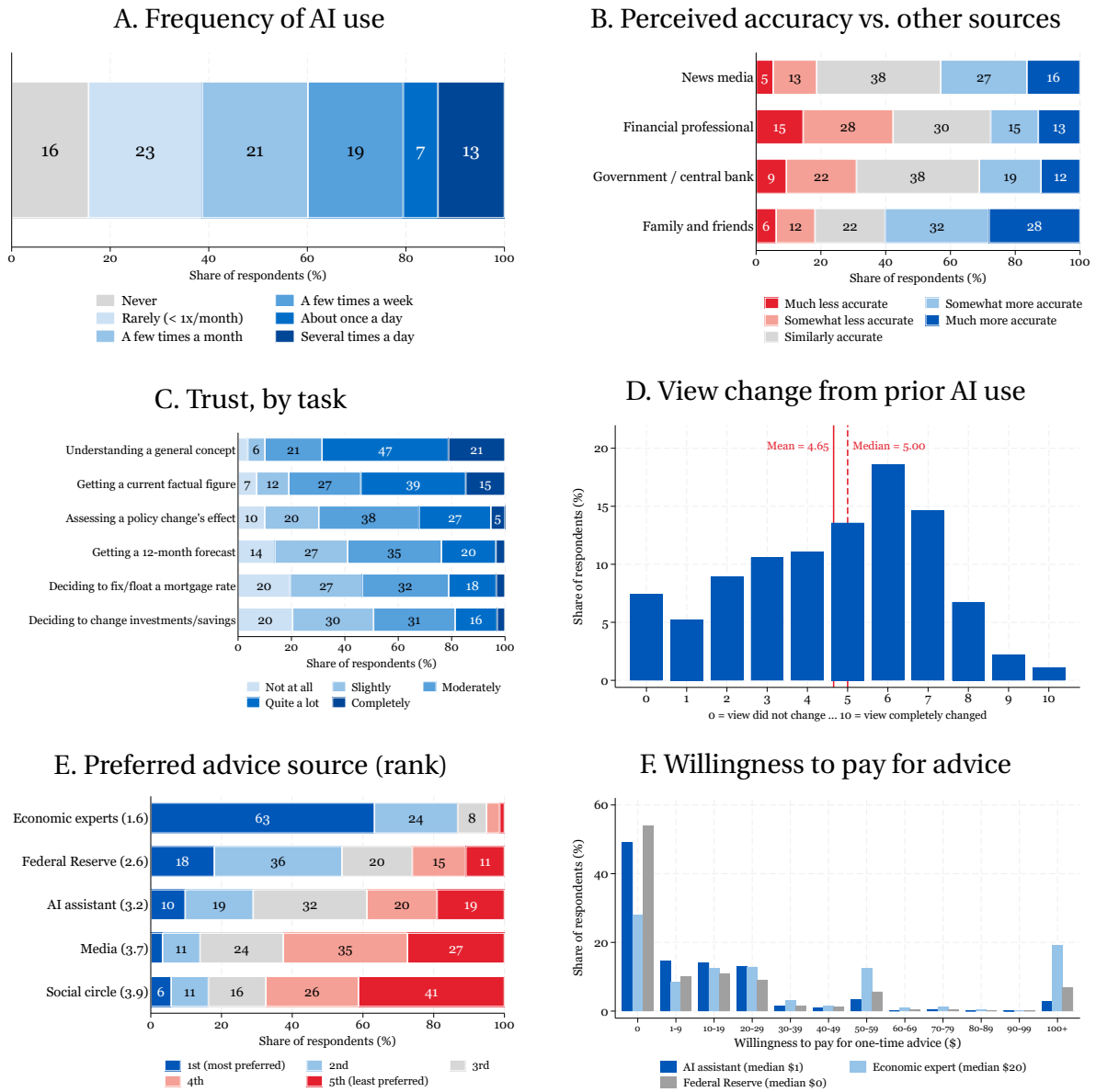
Fact 3: Trust in AI for economic topics varies with the task (Panel C, Figure 4). Trust is highest for understanding a general concept and obtaining a current figure, where 69% and 54% of respondents trust AI "quite a lot" or fully, and lowest for decisions on mortgages and investments (21% and 19%). However, even for these personal decisions, about half of respondents report at least moderate trust.

Fact 4: Respondents report that AI changes their views and expectations on economic topics (Panel D, Figure 4). On a scale from 0 (no change) to 10 (complete change), respondents who regularly use AI for economic topics place the effect of their most recent use at a median of 5, and 57% report a value of 5 or higher.

Fact 5: Households rank AI behind economic experts and the Federal Reserve as a source of advice on the implications of inflation for spending (Panel E, Figure 4). The AI assistant has an average rank of 3.2 among five sources, behind economic experts (1.6) and the Federal Reserve (2.6) but ahead of the media (3.7) and the social circle (3.9); 10% rank it first. Consistent with this relatively favorable view of AI, half of respondents would pay a positive amount for AI advice (Panel F) and 7% would pay $50 or more, against 72% and 34% for expert advice. Given the near-zero costs of AI assistants (at least in free versions), this suggests that households perceive AI assistants as a relevant source of advice for their spending decisions.

Fact 6: The use of AI as a source of macroeconomic information differs systematically across socio-demographic and political groups (Table 1). Employed respondents score 0.13 to 0.32 standard deviations higher on all six (z-scored) measures. Women use AI less often for economic topics (0.10 standard deviations), rate it as less accurate, rank it lower, and would pay less for its advice, but report similar trust. This gap is consistent with the gender gap in generative AI adoption documented in other settings (Otis et al., 2024). It may also have economic consequences: women hold higher inflation expectations (D'Acunto et al., 2021) and score lower on financial literacy (Lusardi and Mitchell, 2008), so whether AI narrows or widens these gaps depends on who uses it and what it advises. Partisanship is associated with the largest differences: relative to Independents, Republicans score between 0.21 and 0.39 standard deviations higher on all six measures. This pattern is in line with recent polling in which Republicans have become less skeptical of AI than Democrats (Pew Research Center, 2026).

Figure 4 AI Usage, Attitudes, and Demand for AI Advice



Notes: All measures are elicited before randomization. $N = 3{,}870$ (3,266 in Panel D, respondents who use AI for economic topics). Panel A: frequency of AI use for economic topics. Panel B: perceived accuracy of AI relative to four other sources (don't-know excluded). Panel C: trust in AI for six tasks. Panel D: extent to which prior AI use changed the respondent's views (0 = not at all, 10 = completely). Panel E: rank of five advice sources (1 = most preferred). Panel F: non-incentivized willingness to pay for advice, in $10 bins; values of $100 or more are pooled.

4 Prior Inflation Expectations, Narratives, and Cognitive Uncertainty

Before turning to the effects of the treatments, we document households' prior beliefs about inflation. We characterize four types of beliefs elicited before randomization: the level of expected inflation, uncertainty about future inflation, the narratives respondents hold about current inflation, and their confidence in translating these expectations into spending decisions.

Table 1 AI Usage and Attitudes on Demographics

	(1) Usage	(2) Trust	(3) Accuracy	(4) View change	(5) WTP	(6) Rank
Female	-0.100***	0.026	-0.061*	0.040	-0.072**	-0.078**
	(0.032)	(0.032)	(0.032)	(0.035)	(0.032)	(0.032)
College degree	0.065*	0.020	0.053	0.059	0.118***	-0.123***
	(0.033)	(0.034)	(0.034)	(0.038)	(0.032)	(0.034)
High income	0.066*	0.144***	0.088***	-0.037	0.062*	0.034
	(0.035)	(0.034)	(0.034)	(0.039)	(0.037)	(0.034)
Employed	0.321***	0.237***	0.132***	0.132***	0.128***	0.158***
	(0.040)	(0.041)	(0.042)	(0.050)	(0.035)	(0.041)
Age 45+	-0.063*	0.160***	0.203***	-0.067*	0.056	0.213***
	(0.034)	(0.033)	(0.033)	(0.038)	(0.036)	(0.034)
Democrat	-0.004	-0.037	-0.115***	-0.034	0.017	-0.190***
	(0.036)	(0.038)	(0.037)	(0.041)	(0.035)	(0.037)
Republican	0.394***	0.327***	0.319***	0.229***	0.270***	0.211***
	(0.043)	(0.041)	(0.041)	(0.046)	(0.048)	(0.043)
Observations	3,870	3,870	3,850	3,266	3,870	3,870
R^2	0.057	0.046	0.050	0.016	0.025	0.048

Notes: OLS. Outcomes are z-scored pre-treatment measures: frequency of AI use for economic topics, trust in AI (mean of six tasks), perceived AI accuracy relative to four other sources, extent AI changed the respondent's economic views, willingness to pay for AI advice, and the reversed rank of the AI assistant among five advice sources. Omitted party category: Independent, other, no answer. Robust standard errors in parentheses. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

To elicit prior inflation expectations, we ask respondents to report a probability distribution over ten fixed bins for inflation over the next twelve months, the same bins used in the New York Fed's Survey of Consumer Expectations (SCE). We calculate each respondent's inflation expectation as the implied mean of this distribution. The median inflation expectation in our sample is 3.55 percent, very close to the 3.63 percent in the July 2026 SCE (Appendix Table A3).[5] The cross-sectional correlates of expectations with household characteristics are consistent with established patterns (Appendix Table A4). For instance, women expect 0.3 percentage points higher inflation than men, consistent with D'Acunto et al. (2021), and employed respondents expect 0.3 percentage points less. The largest gap is partisan, with Republicans on average expecting 0.9 percentage points lower inflation than Independents, almost one-quarter of the sample mean. This pattern is consistent with households forming more optimistic expectations, in this case lower inflation expectations, when their own party holds the White House (Binder et al., 2024). Our survey questions on households' use of AI for macroeconomic information also allow us to relate those to the inflation expectations of respondents: participants who report using AI for economic topics at least a few times a week have 0.47 percentage points lower inflation expectations.

Using the reported subjective probability distributions, we compute the implied standard

[5] Prior forecasts and uncertainty are winsorized at the 1st and 99th percentiles. Appendix Figure A1 plots the average probability placed on each bin and the distribution of implied point forecasts.

deviation, a measure of uncertainty about future inflation. The mean (median) inflation uncertainty of respondents is 2.6 (2.2) percentage points, which is similar to the disagreement across respondents (2.5 percentage points).[6] Uncertainty rises with the level of expected inflation. Uncertainty is lower for high-income and older respondents (Appendix Table A4). Unlike in the euro area (Georgarakos et al., 2024), women in our sample are not more uncertain than men. Republicans, despite expecting lower inflation, report 0.25 percentage points more uncertainty than Independents. Households that use AI intensively for macroeconomic information are also more uncertain, by 0.55 percentage points.

Following the question about future inflation, respondents are asked to name the most important drivers of current inflation, selecting up to two inflation narratives from a list. The listed narratives comprise supply-side causes (tariffs, supply-chain disruptions, energy, labor shortages, climate change), demand-side causes (fiscal and monetary policy), government mismanagement, AI and technology, and an open "Other" field. Respondents' narratives are focused on two drivers of inflation (Appendix Figure A3): government mismanagement (54%) and tariffs (49%). Overall, 78% name at least one supply-side cause, but only 25% name fiscal or monetary policy. Inflation narratives are also highly partisan. Democrats cite tariffs and government mismanagement more often than Republicans (59% vs. 35% and 59% vs. 42%), while Republicans more often blame fiscal or monetary policy for current inflation (32% vs. 22%). Relating inflation narratives to households' expectations, we find that respondents holding the government mismanagement narrative have 0.5 percentage points higher inflation expectations, suggesting that respondents see government mismanagement as a more persistent driver of inflation.

Besides expectations and narratives, we also elicit the prior uncertainty of households about their decisions, that is, how confident they are in their decisions. This information is important in our context because lack of confidence may be a contributing factor to the weak pass-through from beliefs to decisions. Indeed, Enke et al. (forthcoming) show across many decision problems that higher cognitive uncertainty makes choices less elastic to fundamentals. In our context, inflation expectations are the fundamental and household spending is the decision.

Following Enke and Graeber (2023), we measure cognitive uncertainty as respondents' confidence that their spending decisions are optimal given their expectations. Specifically, we ask:

> *How confident would you be to make optimal spending decisions based on your inflation expectations?*

[6] If we calculate uncertainty as in the SCE, as the interquartile range of a fitted distribution, the median in our data of 2.9 percentage points is close to the SCE's 2.7 (Appendix Table A3).

Respondents can answer on a scale from 0 (completely uncertain) to 100 (fully certain). Lower confidence corresponds to higher cognitive uncertainty about the mapping from expectations to decisions.

On average, respondents report a decision confidence of 54, and confidence varies widely across respondents with a standard deviation of 25 points (Appendix Table A3 and Appendix Figure A2). There is systematic variation in the cross section (Appendix Table A5). Women report 7.0 points less confidence than men. Employed, high-income, and college-educated respondents report higher confidence. The partisan gap is comparable to the gender gap in decision confidence: Republicans are 7.2 points more confident than Independents, while Democrats and Independents are similar. Confidence is also correlated with expected inflation. Conditional on demographics and party preference, a one percentage point higher prior forecast is associated with 1.0 point less confidence, and one percentage point more forecast uncertainty with 0.7 points less. Respondents who blame government mismanagement are 4.3 points less confident. Engagement with AI also predicts confidence: each step on the 0–5 AI usage scale adds 1.4 points, and each point of trust in AI advice (on a 0–4 scale) adds 4.2 points. Controlling for AI engagement shrinks the Republican gap to 4.3 points. Thus, confidence in optimal spending decisions is lowest among respondents who expect high and uncertain inflation and attribute it to policy failure.

5 How Households Use our AI Assistant

This section describes how respondents interact with the AI assistant in our survey experiment. We first report the duration and intensity with which households used the AI chatbot. Then we analyze the transcripts to identify which topics respondents were interested in and what the AI assistant's advice looked like. To classify conversations, we use LLMs.

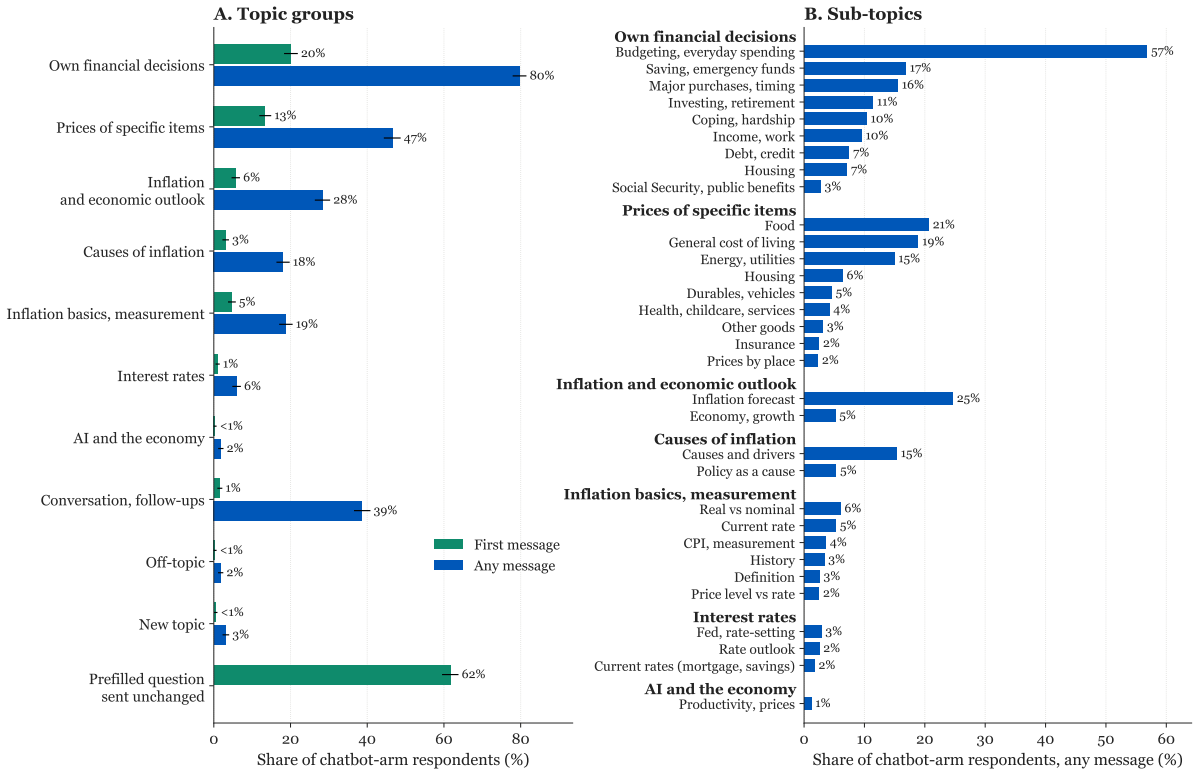
Take-Up of AI assistant Survey respondents in the treatment arms T3 and T4 ($N = 2{,}413$) had the option to interact with an AI assistant for up to five minutes. On average, they spent 4.7 minutes in the conversation, showing strong compliance with the treatment. The median conversation includes four messages from the respondents and four chatbot responses. Engagement does not differ between T3 and T4 (Appendix Figure A4).

Classifying AI conversations With more than 2,000 conversations between respondents and our AI assistant, we can analyze a uniquely rich source of data. With an average conversation length of almost 5 minutes and in total over 10,000 messages, these data are to extensive to code by hand. To analyze the transcripts in a systematic way, we developed an LLM-based classification approach. Before the classification, a research assistant read a random subset of conversations to construct a detailed codebook, which was then reviewed and,

where necessary, corrected by the authors. Based on this codebook, all transcripts were classified using different LLMs via the OpenAI API. We classified household messages and the responses of the AI advice separately. In an independent validation sample, the LLM classification agrees with human coding in 88% of cases.

Household Messages We classified the requests and messages of households into different topics, assigning each message to one or two main topics (Table 2 gives examples). As Figure 5 shows, the vast majority of households (80%) asked the AI assistant about their own financial decisions. Besides financial advice, participants also wanted to know about the prices of particular goods (47%), the outlook for inflation (28%), or how inflation works (19%). 18% wanted to know about the causes of inflation. Only a small share of conversations showed no specific content. Looking into the specific sub-topics respondents were interested in (Panel B in Figure 5), we observe that among those talking about their own financial decisions, a large share asked about budgeting and everyday spending. Among prices of specific items, participants were mostly interested in food, the general cost of living and energy prices. A sizable share of respondents also asked about the inflation forecast (25%).

Figure 5 What Respondents Asked the Chatbot



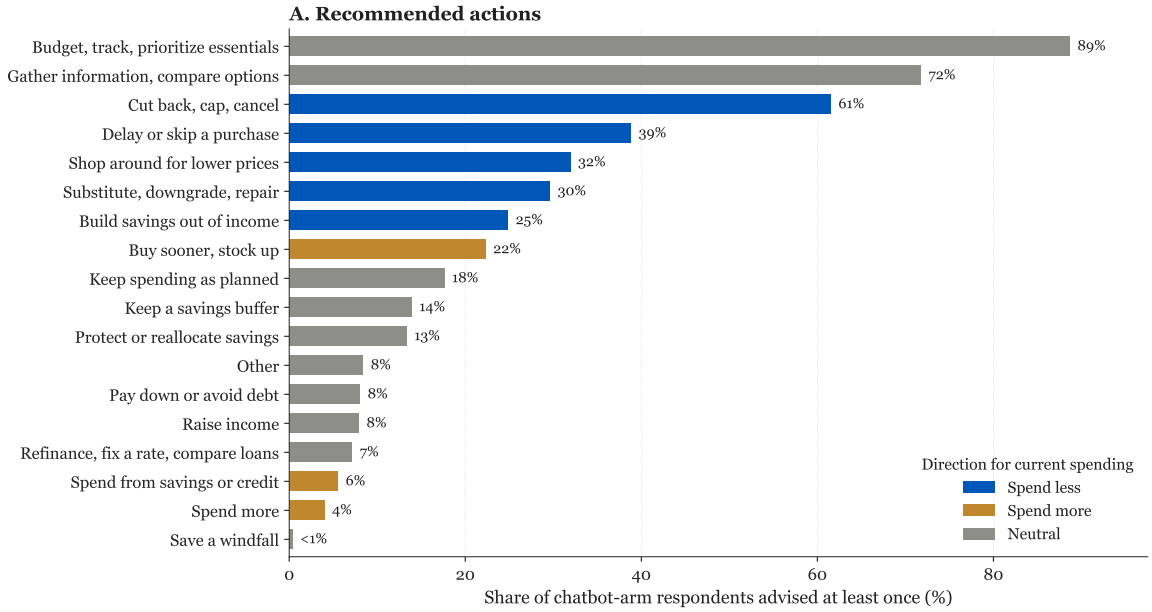
Notes: Topics of respondents' messages to the chatbot, coded by GPT-6.1 Sol, for respondents in the two chatbot arms (N = 2,283). Each message receives a primary and, where substantial, a secondary topic. Panel A: share of respondents whose first message, or any message, touches each topic group; the last row is the share who sent the prefilled question unchanged. Panel B: sub-topics raised by at least 1% of respondents in any message. Bars in Panel A: 95% Wilson confidence intervals.

Table 2 Examples of Questions Respondents Asked the AI Assistant

Type of question	Examples
Own finances (budget, savings, debt, purchases)	
Budgeting and spending	"Where do I cut back if I have already cut back all of the non-essentials?" "What spending categories should my household reduce first during periods of high inflation and which expenses should my household avoid cutting?" "What if my paycheck is already barely covering my expenses?"
Saving and investing	"Should I be putting more of my paycheck into savings or into the stock market? How might inflation affect that decision?" "Is it better to invest in a certificate of deposit account today in order to beat inflation?" "How might inflation affect what I should do to save for retirement?"
Debt and borrowing	"[. . .] would you recommend focusing on paying off debt or saving if I had to choose one goal right now?" "If I put off paying debt, as inflation increases, can delaying payment be a viable strategy?" "Are purchases that involve a loan, such as to buy a car or lease a car, a bad idea at the moment?"
Timing of purchases	"Would it make sense to buy a new car now or wait?" "Can you tell me about big purchases next? I assume that with inflation it's better to buy now before they get even more expensive. Is that the right way to think about it?" "If I need to buy a freezer, should I buy it asap before inflation affects the price?"
Prices of particular goods	"How much will the cost of food go up?" "With oil back down to under $70 is there any chance grocery prices will fall for staples like canned vegetables and frozen potatoes like tater tots?" "Are auto insurance rates affected by this? I feel like those rates keep rising as well and we can't avoid car insurance."
Macroeconomic statistics and forecasts	"What has been the inflation rate this year, 2026?" "What is the estimated inflation rate over the next 12 months?" "Do you think the interest rates will increase or decrease over the next year?" "What's the possibility that this forecast is wrong? What factors could change the equation?"
How and why inflation works	"Inflation has been easing recently. Does that mean prices will start falling, or will they just increase more slowly?" "Why does inflation rise quickly but is slow to decrease?" "What drove this marked increase in inflation since COVID? Was it monetary policy, the COVID bailouts, or something else?" "Who sets the interest rates, is it Warren Buffett?"
AI effects on jobs, wages, and growth	"With AI, automation and now these new self-driving trucks. What are the chances we can fight inflation on the supply side as opposed to raising interest rates like we always do?" "How do I protect my work from being stolen by AI?"
Greetings and chatbot meta-talk	"Who programmed you?" "Are you truly giving accurate inflation information?" "Inflation between 2 and 3 percent sounds low to me. How long ago was your model updated?"
Off-topic or no content	"Let's talk about antinatalism." "What are some opportunities around the US transitioning to a new digital financial system?"

Notes: Respondent messages, lightly edited for spelling and capitalization; [. . .] marks omitted text. The table omits the prefilled opening message.

Figure 6 Actions Recommended by the Chatbot



B. Spending actions by domain (% of respondents)

	General	Food at home	Restau-rants	Leisure	Other non-durables, bills	House-hold durables	Vehicles	Other durables	Housing	Services
Cut back, cap, cancel	40	12	9	32	12	<1	1	<1	1	1
Substitute, downgrade, repair	10	18	3	1	4	<1	1		2	1
Shop around for lower prices	13	12	<1	2	10	2	1	<1	1	1
Delay or skip a purchase	25	<1	<1	9	1	5	4	<1	2	<1
Buy sooner, stock up	14	1		1	1	5	4	1	2	<1
Spend more	1	1	<1	<1	<1	1	1		1	<1
Keep spending as planned	11	3	<1	1	1	1	1		1	<1

Notes: Actions recommended in the chatbot's replies, coded by GPT-6.1 Sol, for respondents in the two chatbot arms (N = 2,283). Panel A: share of respondents advised each action at least once; color gives the action's direction for current spending. Panel B: share of respondents advised each spending action in a given domain.

Content of AI Advice We next describe the advice that the AI assistant gave to respondents. To measure it, we classify the content of the assistant's replies with a large language model. Our codebook included precise rules for coding each sentence of a reply. Every sentence receives a function (for instance, a recommendation or an explanation), and every recommendation receives one or two of sixteen actions (for instance, cutting spending or budgeting more carefully). Recommendations of the seven spending actions also receive the spending domains they target, which mirror the spending categories of our survey. The model applied this codebook to every sentence of all 2,435 conversations, classifying in total 86,742 sentences.[7] Figure 6 reports the recommended actions for the respondents in the two chatbot

[7]We use GPT-6.1 Sol at low reasoning effort; replies whose codes violated the codebook were re-coded, up to three attempts. To assess accuracy, a research assistant coded a subset of conversations at the sentence level. An

arms. The assistant made at least one recommendation in 97 percent of conversations, with a median of four distinct actions. The most common advice concerns how to manage spending rather than how much to spend: 89 percent of respondents were told to budget or prioritize essentials, and 72 percent to gather information. Advice that lowers current spending is far more common than advice that raises it. Cutting back was recommended to 61 percent of respondents, delaying a purchase to 39 percent, shopping around to 32 percent, and substituting to 30 percent. By contrast, buying ahead of price increases was recommended to 22 percent, and spending more for other reasons to only 4 percent. The cuts fall on general and leisure spending, while recommendations about durables reach at most 5 percent of respondents (Panel B of Figure 6). This pattern is consistent with the assistant treating higher inflation as a squeeze on real income rather than as an incentive to bring purchases forward.

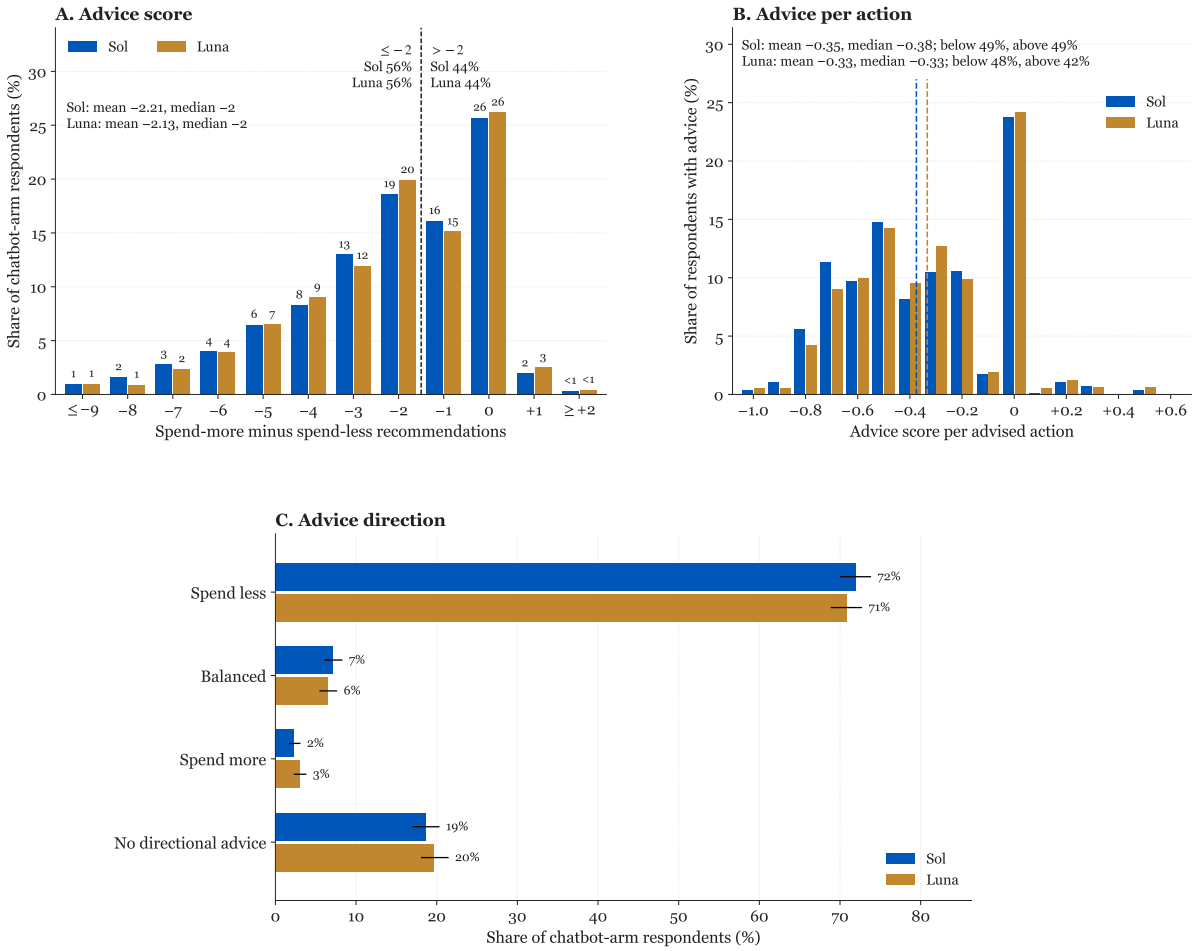
Direction of AI Advice To shed further light on the advice our respondents receive, we construct a measure of spending direction, that is, whether the advice points toward increasing or decreasing spending in response to inflation. Based on the classification of recommended actions described above, we calculate a score of spending advice at the conversation level. We assign each recommended action a direction for current spending: positive (+1) for buying sooner, spending from savings or credit, and spending more; negative (−1) for cutting back, delaying a purchase, substituting, shopping around, and building savings out of income; and neutral (0) for all other actions, such as budgeting or gathering information. We then aggregate these directions into one measure per respondent: each distinct combination of action and spending domain counts once per conversation, and the score equals the number of combinations pointing toward higher spending minus the number pointing toward lower spending. For instance, a respondent who only received advice to cut spending in response to inflation has a negative score, while a respondent who only received budgeting advice has a score of zero. The average score is −2.21 and the median −2 (Panel A of Figure 7), and 72 percent of conversations contain more advice to spend less than to spend more, against 2 percent with the reverse.

Heterogeneity in the Advice of the AI Assistant Because respondents talk to the assistant in their own words, the advice they receive is not assigned by us. To understand where its variation comes from, we first describe how advice accumulates over the conversation, then decompose its variance into its sources, and finally relate it to the content of respondents' messages and to the information treatment.

The assistant's default advice is a modest cut. In its first reply, it recommends, on net, 0.74 more actions to spend less than to spend more (Panel A, Figure 8). The advice becomes more decisive as the conversation continues: recommendations to spend less rise from 0.9 to 2.6,

independent classification by GPT-6 Luna with the same codebook agrees with GPT-6.1 Sol on the net direction of the advice in 90 percent of conversations, and the two advice scores have a correlation of 0.89.

Figure 7 Direction of the Assistant's Spending Advice

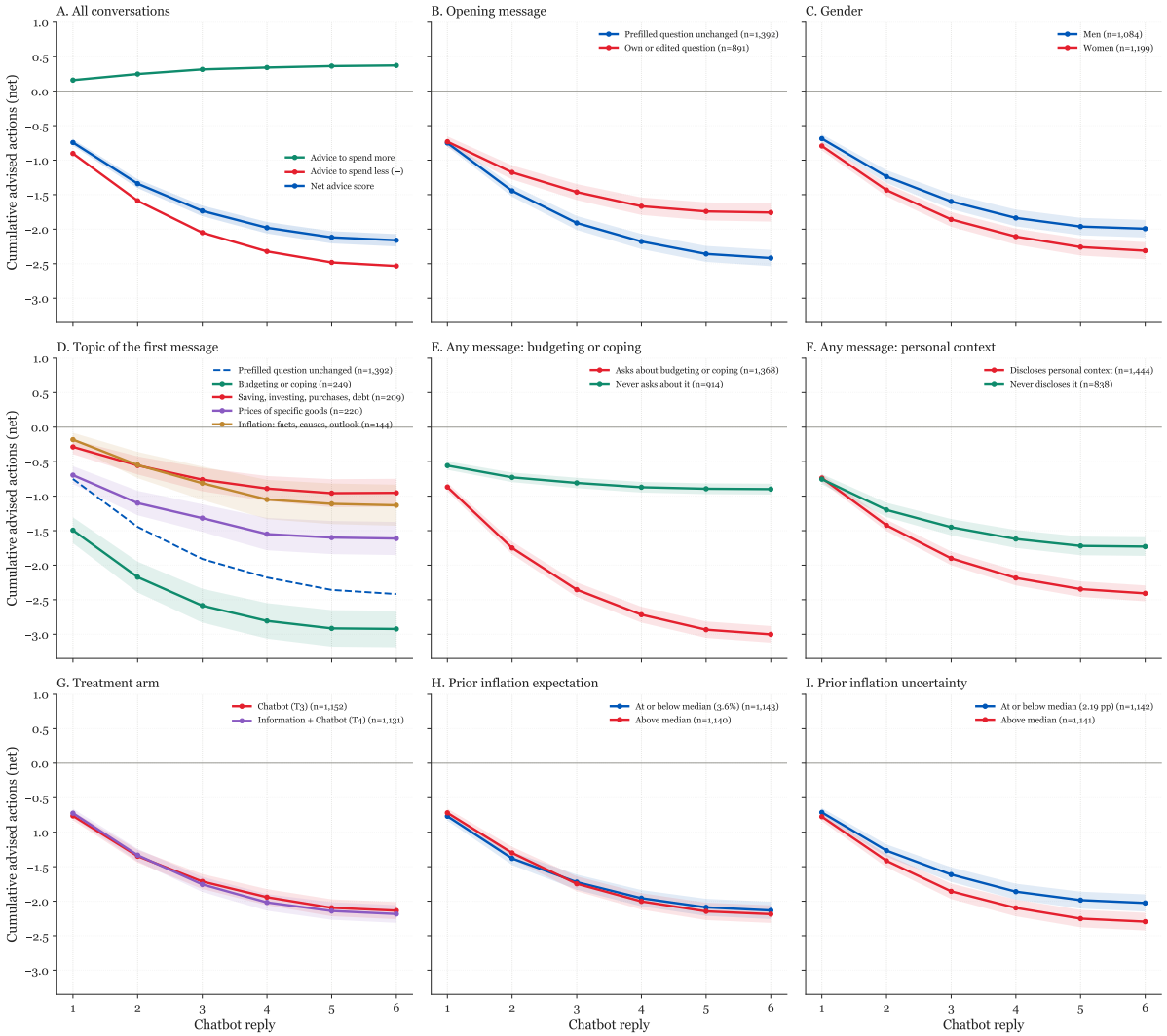


Notes: Spending advice in each chatbot conversation, coded independently by GPT-6.1 Sol and GPT-6 Luna. Each action–domain pair scores +1 (spend more), −1 (spend less), or 0 (neither). Panel A: advice score, the number of spend-more pairs minus the number of spend-less pairs; scores of −9 or below and +2 or above are pooled; the dashed line marks the split at −2. Panel B: advice score divided by all advised pairs, in bins of width 0.1; dashed lines mark each coder's median. Panel C: overall direction of the advice, with 95% Wilson confidence intervals. T3 and T4 pooled. N = 2,283 (Panel B: 2,224 for Sol, 2,222 for Luna).

while recommendations to spend more stay below 0.4, so the net count reaches 2.21 by the end of the conversation. Most of this change occurs in the first four or five replies, beyond which few conversations continue. The first reply is the same whether respondents send the prefilled question or write their own, but the paths then diverge: conversations that start from the prefilled question, which asks directly about household spending, end with more cut-back advice than those that start from respondents' own questions (Panel B).

Respondent characteristics explain little of this variation (Table 3). Women, younger and lower-income respondents, and Republicans receive somewhat more advice to cut back (Panel C, Figure 8; Appendix Table A15), but demographics explain only 2 to 3 percent of the variance in the advice score. Instead, the advice is driven by what respondents ask: their requests account for about one-third of the variance in the whole-conversation advice score. Some of the demographic variation runs through these requests (e.g., women are more likely to ask about budgeting and to state plans to cut expenditures, and older respondents are

Figure 8 Cumulative Spending Advice over the Conversation: Overview



Notes: Mean cumulative advice score (GPT-6.1 Sol coding) after each of the first six chatbot replies; values are carried forward after a conversation ends. Panel D: content of the first message (dashed: prefilled question sent unchanged). Panels E–F: content of any message. Panels H–I: median splits. Bands: 95% confidence intervals.

more likely to write their own question and share personal context; see Panel B, Appendix Table A16). Each reply responds mainly to the current request: earlier requests and the assistant's own earlier advice add little. Requests themselves are only moderately persistent, with a correlation of about 0.3 to 0.4 between the type of a request and the type of the previous one, so the direction of a conversation is not set by its first message.

The type of request shapes the advice: the more respondents reveal about their circumstances, the more the advice tilts toward cutting spending (Panel F, Figure 8). Respondents who share personal context, such as their income or monthly spending, end the conversation with about 2.5 net recommendations to spend less, against about 1.75 for those who never do (Panel F). Those who ask about budgeting or coping end with about 3.1, against about 0.9 for those who never do (Panel E), and those who state a plan to cut back or save in their first message receive the most cut-back advice (Panel D). By contrast, first questions about prices,

Table 3 What Drives the Chatbot's Spending Advice

	(1) Reply 1	(2) Reply 1, prefilled question[b]	(3) Replies 2+	(4) Whole conversation
Current request[a]	12%	–	26%	32%
Topic	3%	–	11%	11%
Request type	6%	–	9%	10%
Stated spending plan	2%	–	5%	6%
Personal context disclosed	1%	–	2%	3%
Prefilled or own question	0%	–	–	1%
Earlier requests	–	–	1%	–
The chatbot's own earlier advice	–	–	2%	–
Demographics shown to the chatbot[c]	0%	0%	0%	2%
Other respondent characteristics[d]	0%	0%	0%	0%
Total explained	11%	0%	30%	34%
Unexplained[e]	89%	100%	70%	66%
Replies	2,282	1,392	7,668	2,282
Respondents	2,282	1,392	2,195	2,282

Notes: Shapley decomposition of the cross-validated R^2 of the advice score (actions advising more minus less spending). [a]Column 4: requests in any message. [b]Respondents who sent the prefilled question unchanged. [c]Age, gender, education, income, employment status, and monthly non-durable spending, which the scripted opening message passes to the chatbot. [d]Party and prior AI use, which the chatbot does not see. [e]Remainder, including measurement error in the LLM-coded score.

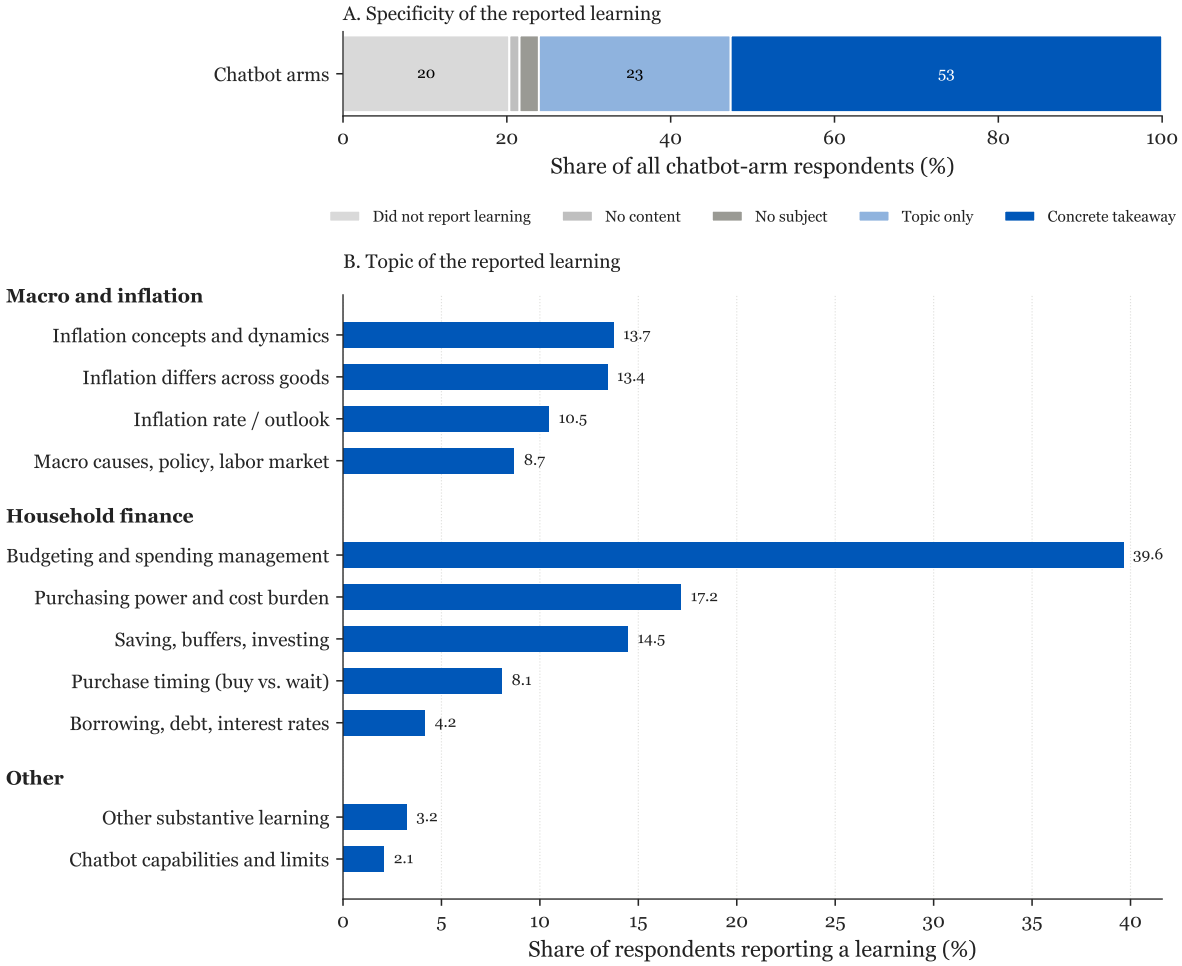
the inflation outlook, or saving, investing, purchases, and debt lead to less cut-back advice (Panel D).

One concern is that the requests might themselves be shaped by the information treatment: respondents who saw the inflation statistic and revised their expectations could ask for different advice. The data do not support this concern (Panel G, Figure 8). The distribution of advice scores is similar in the two arms, the arms accumulate advice along nearly identical paths, and the shares of conversations by direction of advice do not differ (Appendix Figure A5). The advice also does not vary with respondents' prior inflation expectations (Panel H, Figure 8). Hence, the advice is not itself a consequence of the information treatment.

Taken together, the systematic component of the advice largely reflects what respondents ask and how much they reveal, while about two-thirds of the variance in the whole-conversation advice score remains unexplained (column (4), Table 3). Among the part of the variation in spending advice explained by respondents' requests, stated plans are only a small share, while a considerable share can be explained by the topic (e.g., whether someone talks about budgeting or about the causes of inflation) or the type of request. We therefore conclude that the direction of advice that respondents receive is quasi-random, as it unlikely reflects demographic sorting or seems unaffected by the information treatment. For respondents who sent the prefilled question unchanged (61 percent of chatbot-arm respondents), the first reply is essentially unrelated to their characteristics, so differences in their initial advice are close to random.

Rating of Experience and Learnings Right after the AI assistant intervention, we asked respondents how helpful they found the previous AI conversation, and on average they find it fairly helpful (6.8 on a scale from 1 to 10). In addition, we asked them whether they learned something and, if so, what they learned (Figure 9). The vast majority (80%) indicate that they learned something. Among those, a large share cites a concrete takeaway from the AI conversation, and a considerable part mentions having learned about a topic. In terms of topics, household finance dominates, with many respondents citing budgeting and spending management (40% of respondents reporting to have learned something), purchasing power and cost burden (17%) and saving and investing (15%). To a lesser extent, participants also reported takeaways for the macroeconomy and inflation: inflation concepts and dynamics (14%), inflation differences across goods (13%) and the inflation rate or its outlook (11%).

Figure 9 Self-Reported Learning from the Chatbot Conversation



Notes: Open-ended answers to "What did you learn from the interaction?", asked of chatbot-arm respondents who reported learning something, coded by GPT-6 Luna. Panel A: specificity of the answer, one code per respondent (N = 2,413, including respondents who reported no learning). Panel B: topics of the reported learning; one answer can carry several topics (N = 1,923).

6 RCT Results

So far we have described observational evidence on households' views on AI as well as AI usage. In the next step, we exploit the randomized variation in the provision of information and access to an AI chatbot to shed more light on how these interventions shape beliefs and actions.

To measure the effects of treatments on beliefs, we use a standard specification (Coibion and Gorodnichenko 2026):

$$x_i^{post} = a_0 + b_0\, x_i^{prior} + \sum_{j=2}^{4} \gamma_j \times I\{i \in Treat\ j\} + \sum_{j=2}^{4} \delta_j \times I\{i \in Treat\ j\} \times x_i^{prior} + error_i, \quad (1)$$

where x^{prior} and x^{post} are prior and posterior beliefs, $I\{i \in Treat\ j\}$ is an indicator variable equal to one if respondent i is in arm j (T2: information, T3: chatbot, T4: information and chatbot), with the control arm (T1) being the omitted category. Coefficients γ and δ measure "level" and "slope" effects, that is, how beliefs are shifted on average and how strongly they are updated towards signals provided in the treatments. With Bayesian updating, the "slope" effects should be between 0 (no effect) and -1 (priors are abandoned and beliefs are completely shaped by the signal in the treatment). One should expect b_0 to be close to one (posteriors should equal priors in the control group, which receives no information), but because we elicit prior and posterior beliefs with different questions,[8] b_0 could deviate from unit slope and so one can use $\hat{b}_0$ to normalize δ when measuring departures from FIRE (see Weber et al. 2025 for a discussion). Because we rely on randomization to create variation in treatments, including controls such as age and gender does not bias the estimates but can improve their precision by reducing the variance of the error term.

In addition, we estimate the following specification that gives us the *average* treatment effects (ATEs) on beliefs:

$$x_i^{post} = a_0 + \sum_{j=2}^{4} \psi_j \times I\{i \in Treat\ j\} + error_i, \quad (2)$$

where ψ_2, ψ_3, ψ_4 measure ATEs for T2, T3 and T4 respectively. To be clear, specification (1) utilizes more information, but specification (2) provides useful summary statistics.[9] To reduce adverse effects of outliers and influential observations, we estimate both specifications

[8] For example, for priors, we ask respondents to assign probabilities for 10 pre-set scenarios ("bins"); for posteriors we ask respondents to give us their point forecast as well as define three inflation scenarios (low, medium, high) and assign probabilities to these.

[9] To see the benefit of specification (1), suppose that half of respondents expect inflation to be 2 percent and the other half expect 6 percent. Also suppose that after an information treatment (say, inform respondents that past inflation is 4 percent), posterior beliefs become 3 and 5 percent respectively. While the average effect (captured by ψ in specification (2)) is zero (the average prior is 4 percent and the average posterior is 4 percent), coefficient δ in specification (1) captures that posterior beliefs become closer to the provided signal.

using Huber-robust regressions. We report results for inflation expectations and cognitive uncertainty in Tables 4 and 5 and visualize the results in Appendix Figures A6 and A7.

In agreement with earlier evidence (Armantier et al., 2016; Cavallo et al., 2017; Binder and Rodrigue, 2018; Coibion et al., 2022), we observe that informing respondents about past inflation (group T2) materially affects their inflation expectations (column (1) of Table 4), which is not consistent with full-information rational expectations (FIRE) because this information is publicly available. Having a conversation with AI (group T3) has a tangibly weaker effect on inflation expectations: the slope coefficient δ for T3 is about half of that for T2, and the level coefficient γ about one-third. This pattern is consistent with AI conversations being generally focused on concepts and practical advice rather than macroeconomic statistics or forecasts. Providing respondents with inflation information and access to AI (T4) yields results broadly similar to the estimates for T2. Column (2) of Table 4 shows that treatment effects are present in the follow-up wave and thus changes in beliefs shown in column (1) are unlikely to be driven by anchoring or similar effects.[10] In terms of ATEs (column (1) in Table 5), T2 and T4 raise inflation expectations by 0.7 and 0.4 percentage points, respectively, while T3 does not materially change inflation expectations.[11]

In terms of inflation uncertainty, we find that all treatments have broadly similar effects and persistence (columns 3 and 4, Table 4). We also observe that the power of interventions is considerably weaker: the F-statistic for treatments and their interactions with priors (which are excluded instruments in later analysis) falls by an order of magnitude from ≈ 100 for point forecasts of inflation to ≈ 10 for the second moment of inflation expectations. Column (2) in Table 5 conveys a similar message: the ATEs are broadly similar on average but treatments account for a relatively small share of variation in posterior beliefs.

The quantitative effects are even weaker for decision confidence for purchases of durable goods or non-durables (columns 5 and 6 in Table 4), with the corresponding F-statistics falling further. The average treatment effects (columns (3) and (4), Table 5) are generally insignificant with the exception of treatment T4, which significantly lowers cognitive confidence by 2 points for purchases of durable goods.

These results suggest that information treatments are mainly working via altering the first

[10] The chatbot conversation also has a lasting effect on attitudes toward AI (Appendix Table A7). Holding main-survey trust fixed, follow-up trust in AI (on a 0–4 scale) is 0.084 points higher in T3 and 0.056 points higher in T4, and perceived AI accuracy rises by 0.099 and 0.066; the information screen alone has no effect. Appendix Tables A7–A9 report additional results on the longer-term effects of exposure to AI in our survey.

[11] We also examine heterogeneity in belief updating by splitting the sample along socioeconomic characteristics and re-estimating column (1) of Table 4 for each group (Appendix Figure A9). Belief updating is quite similar across groups. Point estimates suggest somewhat weaker updating among respondents with high comprehension of inflation and somewhat stronger updating among respondents with low confidence in their durable-spending decisions, but these differences are imprecisely estimated. Notably, Republicans and non-Republicans update by the same amount, so the partisan differences in pass-through discussed below do not reflect differences in how strongly beliefs respond to the information treatment.

moment (i.e., mean) of inflation expectations. Other channels (uncertainty in the inflation outlook and cognitive uncertainty about the spending decisions) are potentially present but they tend to be considerably weaker. This means that treatment-induced changes in inflation uncertainty and cognitive uncertainty are unlikely to drive quantitatively important differences in behavioral responses.

Table 4 Belief Updating in the Main Survey and the Follow-Up

	Inflation expectations				Decision confidence		
	Point forecast		Uncertainty		Durables	Non-durables	General
	Main	Follow-up	Main	Follow-up	Main	Main	Follow-up
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
T2 (Information)	2.397***	1.109***	0.373***	0.364***	5.263**	1.389	5.598**
	(0.123)	(0.178)	(0.062)	(0.062)	(2.469)	(2.603)	(2.546)
T3 (Chatbot)	0.816***	0.870***	0.257***	0.360***	4.841**	3.499	5.244**
	(0.116)	(0.158)	(0.055)	(0.054)	(2.191)	(2.334)	(2.339)
T4 (Info + Chatbot)	2.025***	1.165***	0.296***	0.337***	6.033***	6.019***	4.779**
	(0.110)	(0.159)	(0.056)	(0.055)	(2.177)	(2.329)	(2.249)
Prior	0.772***	0.479***	0.349***	0.280***	0.579***	0.583***	0.603***
	(0.023)	(0.036)	(0.021)	(0.021)	(0.029)	(0.030)	(0.028)
Prior × T2	-0.475***	-0.294***	-0.204***	-0.154***	-0.092**	-0.027	-0.113***
	(0.031)	(0.046)	(0.027)	(0.027)	(0.042)	(0.043)	(0.041)
Prior × T3	-0.230***	-0.242***	-0.153***	-0.163***	-0.076**	-0.062	-0.096**
	(0.030)	(0.041)	(0.025)	(0.024)	(0.038)	(0.039)	(0.038)
Prior × T4	-0.451***	-0.279***	-0.165***	-0.152***	-0.129***	-0.114***	-0.099***
	(0.028)	(0.042)	(0.025)	(0.024)	(0.038)	(0.039)	(0.037)
Control group mean	4.147	3.651	1.348	1.137	48.289	50.927	55.780
Observations	3,648	2,883	3,701	2,912	3,870	3,870	3,074
R^2	0.499	0.195	0.235	0.165	0.256	0.270	0.329
F-test (T2–T4, Prior × T2–T4)	105.27	10.79	10.79	9.15	2.58	1.62	1.94
p-value	<0.001	<0.001	<0.001	<0.001	0.017	0.138	0.071
Equality tests, p-values							
T2 = T3	<0.001	0.084	0.026	0.932	0.856	0.378	0.884
T2 = T4	<0.001	0.686	0.144	0.603	0.740	0.053	0.726
T3 = T4	<0.001	0.009	0.378	0.593	0.554	0.228	0.825
Prior × T2 = Prior × T3	<0.001	0.141	0.019	0.678	0.682	0.379	0.654
Prior × T2 = Prior × T4	0.371	0.684	0.068	0.910	0.352	0.031	0.717
Prior × T3 = Prior × T4	<0.001	0.194	0.536	0.526	0.125	0.145	0.912

Notes: Estimates of equation (1) for main-survey and follow-up outcomes, using the wave-1 prior throughout. Columns (1)–(2): inflation forecast. Columns (3)–(4): forecast uncertainty (percentage points). Columns (5)–(7): decision confidence; the follow-up asks one question, without a durable/non-durable split. Follow-up columns use follow-up respondents only. Robust standard errors in parentheses. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

7 Expectations-to-decisions pass-through

To measure pass-through from expectations to actions, we use the popular two-stage specification (Coibion and Gorodnichenko 2026) where the second stage is given by

$$Outcome_i = \beta\, Post_i^{mean} + \alpha\, Prior_i^{mean} + \xi\, Plan_i + Controls_i + error_i, \tag{3}$$

where $Outcome_i$ is (planned) spending, portfolio allocation, etc. of respondent i, $Post_i^{mean}$ is the posterior (immediately after treatment) twelve-month-ahead inflation forecast, $Prior_i^{mean}$ is the mean of the pre-treatment density forecast, $Plan_i$ is the pre-treatment value (or plan)

Table 5 Posterior Beliefs by Treatment Arm

	Inflation expectations		Decision confidence	
	Point forecast	Uncertainty	Durables	Non-durables
	(1)	(2)	(3)	(4)
T2 (Information)	0.723***	-0.056*	-0.348	-1.117
	(0.058)	(0.031)	(1.241)	(1.277)
T3 (Chatbot)	-0.069	-0.064**	0.647	-0.409
	(0.053)	(0.029)	(1.114)	(1.130)
T4 (Info + Chatbot)	0.412***	-0.037	-2.027*	-1.723
	(0.052)	(0.028)	(1.108)	(1.120)
Control group mean	3.891	1.311	48.289	50.927
Observations	3,570	3,701	3,870	3,870
R^2	0.064	0.001	0.002	0.001
F-test (T2–T4)	98.94	1.84	2.74	1.01
p-value	<0.001	0.137	0.042	0.387

Notes: Each posterior is regressed on arm dummies T2–T4 only (T1 omitted; no prior or other controls). Outcomes and sample as in Table 4. The F-test is the joint test of T2–T4. Robust standard errors in parentheses. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

of the outcome when available, and $Controls_i$ is a vector of respondent characteristics. The first stage is given by a version of equation (1):

$$\begin{aligned} Post_i^{mean} = a_0 + b_0\, Prior_i^{mean} + \sum_{j=2}^{4} a_j \times I\{i \in Treat\ j\} \\ + \sum_{j=2}^{4} b_j \times I\{i \in Treat\ j\} \times Prior_i^{mean} + \kappa\, Plan_i + Controls_i + error_i. \end{aligned} \tag{4}$$

The first stage is estimated by Huber regression, which downweighs outliers, and a jackknife procedure removes influential observations in the second stage.

The pass-through is measured by coefficient β in equation (3). Because 2SLS isolates exogenous variation in $Post_i^{mean}$ due to randomized interventions, β has a causal interpretation. We note that β estimates the *total* effect of inflation expectations on a given outcome variable. In other words, β measures a combination of the direct effect (i.e., how action responds to an exogenous change in inflation expectations holding everything else constant) and indirect effects (e.g., responses to updates of other expectations). As a result, β is not a fixed, structural parameter and the magnitude of β can depend on a variety of factors such as the macroeconomic environment (e.g., Weber et al. 2025, Candia 2026), political views (e.g., Kamdar and Ray 2024), etc. Our thesis is that β can also be affected by how respondents think about the mapping from expectations to actions.

7.1 Baseline

To provide a benchmark, we estimate the pass-through for the full sample (with all treatments) and report the results in Table 6. When we pool across treatments, the first stage is generally

strong: F-statistic is close to 90. We find that a one percentage point increase in expected inflation raises spending on durable goods by 0.2 log points, or roughly 20%.[12] This response is large, but there are several reasons why it is plausible. First, note that the regressand is the log of one plus planned spending and so the estimated slope combines the intensive margin with the extensive margin of whether to plan a durable purchase at all. Moving from no planned purchase to a typical purchase changes the outcome by several log points, so an increase of a few percentage points in the probability of planning a purchase is enough to generate a coefficient of this size. Previous work found that the extensive margin of large purchases responds strongly to households' macroeconomic beliefs (e.g., Coibion et al., 2024), and durable spending can respond to higher inflation expectations within a month (Kostyshyna and Petersen, 2024). Second, our survey took place in an unusually uncertain environment: the Gulf war that began in February 2026 had disrupted oil shipments through the Strait of Hormuz, and headline inflation stood at 4.2 percent, a few years after the largest inflation surge in four decades. These developments can make households particularly attentive and sensitive to inflation (Korenok et al., 2026; Pfäuti, forthcoming). Third, when inflation is high, households respond to exogenously higher inflation expectations by bringing forward durable purchases (Candia, 2026).

Table 6 2SLS: Posterior Inflation Forecast on Planned Spending Decisions

	(1) Log durable sp.	(2) Log nondur. sp.	(3) Risky share	(4) Cons. index (z)
Posterior inflation forecast	0.191**	0.102	-3.8***	-0.020
	(0.084)	(0.064)	(1.2)	(0.028)
Durable-purchase plan control	Yes	No	No	No
First-stage F (robust)	89.2	86.7	77.4	88.0
Observations	3,504	3,503	3,430	3,506

Notes: 2SLS with a Huber-robust first stage and a jackknife trim. All regressions control for the prior forecast, gender, college degree, household income, employment, and age; column (1) adds the pre-treatment durable-purchase plan. Log durable sp. is log(1 + planned durable-goods spending over the next 12 months); log nondur. sp. is log(1 + planned monthly non-durable spending). Risky share is the share (0–100) of a hypothetical $10,000 windfall allocated to stocks, mutual funds, and crypto assets. Cons. index (z) is the mean of five standardized items on the expected change in July relative to June in consumption of durable goods, food at home, restaurants, non-durable goods, and leisure (−2 to +2). Robust standard errors in parentheses. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

The point estimate for non-durable spending (0.10) is not statistically significant, and its confidence interval includes the small responses implied by standard estimates of intertemporal substitution (Crump et al., 2022). Part of any response is also mechanical: because

[12] We also examine heterogeneity in the pass-through by splitting the sample along various characteristics (demographics, political affiliation, treatment arm, AI usage and attitudes, literacy, decision confidence, and inflation narratives) and re-estimating the durable-spending specification of Table 6 for each group (Appendix Figures A10–A11). The main limitation is statistical power: confidence intervals are wide and mostly overlap with each other. Some point estimates differ, for instance larger responses among respondents aged 45 or older and among frequent AI users, but we view these split-sample gaps as suggestive rather than conclusive. The clearest difference is partisan. A 1 p.p. increase in expected inflation raises planned durable spending by 0.20 log points among non-Republicans but has no significant effect among Republicans.

planned spending is reported in dollars, a one percentage point increase in expected inflation raises planned nominal spending by up to 1 percent even if planned real consumption is unchanged. Earlier studies likewise find weak or imprecisely estimated responses of non-durable spending (Burke and Ozdagli, 2023; Coibion et al., 2023).

When we ask respondents to allocate a hypothetical $10,000 windfall across various asset classes,[13] respondents with exogenously higher inflation expectations put a smaller share of their portfolio in risky assets (mainly due to lower investments in stocks and mutual funds, see Appendix Table A6), which is broadly consistent with earlier studies (e.g., Georgarakos et al. 2024; Schnorpfeil et al. 2025). When we use qualitative responses (increase a lot, increase somewhat, etc.) to construct an alternative measure of the spending response (column (4), Table 6), we do not find any significant effect, which underscores the importance of eliciting quantitative responses for consumer spending.[14]

In the next step, we estimate pass-through for one treatment at a time or various subsets of respondents within a treatment or a group of treatments. The main objective of this exercise is to establish how β varies depending on whether and which advice was given to a respondent. We report results in columns (1) and (2) of Table 7. In Panel A, we restrict the sample to the control group and T2 group (information treatment with information about past inflation) so that the estimated pass-through β is not colored by advice from AI. We find that spending on both durables and non-durables (columns 1 and 2) increases in response to higher inflation expectations, with a significant effect for durables. We also find that portfolio allocations to risky assets are not sensitive to higher inflation expectations.

In Panel B, we focus on T1 vs T3. Because a conversation with AI has a much smaller effect on inflation expectations, the first stage is weak for this pair (F-statistics are ≈ 5) and so estimated pass-throughs are less precise and we therefore interpret them with caution.

In Panel C, we restrict the sample to the control group and T4 group (information treatment with information about past inflation *and* access to AI). Now the estimated β captures the effect of both having different inflation expectations *and* having advice from AI. As we documented earlier, AI generally advises respondents to reduce spending if expected inflation

[13] Households are asked how much they would invest in each of the following asset classes: savings or current accounts, stocks and shares, mutual funds, retirement or pension products, bonds, gold or other precious metals, or bitcoin and other crypto assets. We classify stocks, mutual funds, and crypto assets as risky.

[14] Appendix Figure A8 reports the reduced-form estimates. Because the information treatment effect should scale with the dosage (i.e., the difference between the signal and the prior; Haaland et al., 2023), we estimate each arm's effect nonparametrically within bins of this gap. The patterns are visible but noisy. For durables, the slope is positive for information alone (T2), flatter with the chatbot (T4), and slightly negative for the chatbot alone (T3). While reduced-form estimates are informative, we prefer 2SLS for several reasons. First, β in equation (3) measures the response of spending to expected inflation rather than to a particular message, so it is more comparable across arms, studies, and theory. Second, scaling by the first stage separates pass-through from how strongly each arm moves beliefs: a flat reduced form in an arm that barely moves beliefs says little about pass-through. Third, imposing that all experimental variation works through beliefs pools it into one, more precise estimate.

is high. Consistent with this advice, respondents exhibit a lower elasticity of spending with respect to inflation. In the case of durable spending, the difference is statistically significant. Furthermore, respondents now allocate a significantly lower share to risky assets.

To further examine the importance of AI advice, we split T4 respondents by the direction of the advice they received. Panel D includes respondents whose advice clearly points toward cutting spending: it contains at least two more spend-less than spend-more recommendations (a score of -2 or below, which covers 56% of chatbot-arm respondents).[15] Panel E includes the remaining respondents, whose advice does not clearly point toward cutting expenditures. The estimated durable-spending pass-throughs for these groups are -0.194 and 0.388 respectively. This difference is economically and statistically significant. For other planned actions, we do not find statistically significant differences, likely due to lower statistical power.

In Panels F and G, we add respondents with strong and weak AI advice from group T3, who revise their inflation expectations by a smaller extent. The qualitative results are similar to Panels D and E. This pattern suggests that having advice in itself is not enough to spur changes in spending or portfolio choices. That is, one needs both changes in inflation expectations and guidance on the course of actions to overturn the baseline response (Panel A).

Because we elicit actual spending in the follow-up wave, we can check whether differences in spending plans translated to differential spending. We find that plans are strong, but imperfect predictors of actual spending (Appendix Table A10). Part of the gap reflects timing: plans cover the next twelve months, whereas realized spending covers July, the month of the main survey, so purchases planned for later in the year cannot yet show up. In addition, about 21% of respondents did not complete the follow-up, which reduces precision. With these caveats in mind, we re-estimate the pass-through for realized spending (columns (4) and (5) of Table 7). By and large, the role of AI advice carries over. For example, among respondents who received strong advice to cut spending, higher expected inflation does not raise realized spending, whereas among those who received weaker advice it raises realized durable and non-durable spending; the difference between the two groups is statistically significant and at least as large as for planned spending. Because realized spending is reported a month after the advice, and before respondents talk to the assistant again in the follow-up, the advice-related differences in planned spending are unlikely to merely reflect respondents restating the advice they had just received.

The AI conversation also shapes how respondents allocate a hypothetical windfall of $10,000 across asset classes (column (3) of Table 7), but differently than it shapes spending. Without

[15] For sample splits, we use advice scores from Panel A of Figure 7 to capture the totality of AI guidance (heuristically, intensive and extensive margins). Appendix Table A14 reports estimates based on the splits by scores that capture the consistency of advice per action (Panel B, Figure 7). The results are similar.

the conversation, the share allocated to risky assets does not respond to expected inflation (Panel A). With the conversation, a one percentage point increase in expected inflation lowers the risky share by 3.3 percentage points (Panel C), and the difference between the two is significant at the 10% level. Unlike the spending response, however, the portfolio response does not depend on the direction of the spending advice: the coefficient is the same for respondents whose advice score is −2 or below and for those above −2 (Panels D and E). A natural explanation is that the portfolio response reflects a recommendation that is common across conversations, regardless of the direction of the spending advice.

7.2 Examining the mechanism

Our results point to the potential importance of guidance for shaping behavioral responses to exogenous variation in expectations. As we discussed earlier, our estimates provide the total effect that summarizes responses along many margins. While one can disentangle channels with additional treatments targeting specific channels,[16] our research design has a limited number of interventions and thus we take a different route to understand how behavioral responses may be affected by information treatments activating other beliefs. In a nutshell, we examine whether the treatments affect beliefs along a specific dimension. If they do not materially change those beliefs, that dimension is unlikely to drive our results.

7.2.1 Narratives

One potentially confounding factor is that respondents can change views on the sources of inflation or the nature of government response. To assess this hypothesis, we ask respondents to report their (posterior) views on the matter and then estimate specification (2) to investigate whether the treatments are systematically related to these views. Tables 8 and 9 show that, in general, there is little systematic variation. For example, treatments do not cause people to have different views on the role of tariffs in determining inflation (row 5 in Panel B, Table 8). Treatments also do not affect the supply-demand mix in the sources of inflation (Panel A, Table 8). The main exception is that people exposed to AI conversations in T3 are less likely to report government mismanagement (the most popular explanation of inflation) as a driver of inflation. In terms of the desired government response (Table 9), treatments only systematically reduce the appeal of lower taxes, tariffs, and regulation. We interpret these findings as suggesting that, while treatments can move a handful of views, the worldview is largely unaffected by the treatments. As a result, it is unlikely that respondents reduce consumer spending in response to higher inflation expectations after a conversation with AI because they update their beliefs about the drivers of inflation or the desired government response.

[16] For example, Georgarakos et al. (2024) use information treatments separately targeting first and second moments of inflation expectations to create differential variation in the level vs. uncertainty in inflation outlook.

Table 7 Belief Pass-Through into Spending by Chatbot Advice

	Log planned spending		Risky	Log realized spending	
	Durable	Nondurable	share	Durable	Nondurable
	(1)	(2)	(3)	(4)	(5)
Panel A: T1 and T2 (Information)					
Posterior inflation forecast	0.212**	0.116*	0.3	-0.118	-0.103
	(0.098)	(0.068)	(1.5)	(0.101)	(0.064)
First-stage F	141.3	150.3	109.0	121.5	124.5
Observations	1,315	1,321	1,218	1,040	1,051
Panel B: T1 and T3 (Chatbot)					
Posterior inflation forecast	0.018	-0.618**	-8.2	0.779***	-0.079
	(0.375)	(0.301)	(7.3)	(0.270)	(0.156)
First-stage F	6.7	5.9	3.1	14.6	12.2
Observations	1,733	1,731	1,656	1,371	1,366
Panel C: T1 and T4 (Information + Chatbot)					
Posterior inflation forecast	-0.023	0.083	-0.9	-0.065	-0.063
	(0.121)	(0.083)	(1.7)	(0.126)	(0.071)
First-stage F	97.0	93.5	83.3	81.1	79.9
Observations	1,733	1,736	1,647	1,377	1,375
Panel D: T1 and T4, advice score ≤ -2					
Posterior inflation forecast	-0.194*	0.111	-2.3	-0.081	-0.117*
	(0.110)	(0.083)	(1.8)	(0.109)	(0.067)
First-stage F	95.4	91.1	68.3	78.4	82.7
Observations	1,252	1,254	1,159	971	980
Panel E: T1 and T4, advice score > -2					
Posterior inflation forecast	0.388**	0.261*	-3.1	0.048	0.127
	(0.191)	(0.134)	(3.2)	(0.201)	(0.102)
First-stage F	36.9	36.9	29.1	31.6	37.5
Observations	1,109	1,110	1,024	880	890
Panel F: T1, T3, and T4, advice score ≤ -2					
Posterior inflation forecast	-0.143	0.101	-3.4*	0.009	-0.107*
	(0.103)	(0.075)	(1.8)	(0.104)	(0.062)
First-stage F	58.5	54.2	36.0	44.9	43.8
Observations	1,857	1,853	1,770	1,451	1,453
Panel G: T1, T3, and T4, advice score > -2					
Posterior inflation forecast	0.297*	-0.010	-6.5***	0.434***	0.111
	(0.158)	(0.110)	(2.5)	(0.167)	(0.086)
First-stage F	32.8	33.0	28.1	30.8	32.2
Observations	1,596	1,602	1,520	1,278	1,275
Panel H: Equality of posterior-forecast coefficients (Wald p-values)					
A = C	0.038**	0.674	0.431	0.648	0.567
D = E	0.001***	0.232	0.804	0.470	0.010***
A = D	0.000***	0.958	0.140	0.732	0.840
A = F	0.003***	0.854	0.070*	0.276	0.953
A = G	0.600	0.254	0.010***	0.002***	0.019**
F = G	0.008***	0.348	0.272	0.014**	0.016**
Pre-treatment spending control	Yes	Yes	No	Yes	Yes

Notes: 2SLS as in Table 6. Columns (1)–(3): planned log(1+durable spending), planned log(1+non-durable spending), and the share of windfall $10,000 allocated to risky assets (risky share). Columns (4)–(5): realized log(1+ durable) and log(1+non-durable spending). Spending columns add their own pre-treatment control: an indicator for a planned durable purchase (durable) and log baseline non-durable spending (non-durable). Panels D–G split T3 and T4 respondents at a GPT-6.1 Sol advice score of −2 (Panel A of Figure 7). Panel H: Wald p-values. Robust standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$ (on the coefficient in Panels A–G, on the equality test in Panel H).

Table 8 Treatment Effects on Posterior Inflation Narratives

	Control mean	T2 (Information)	T3 (Chatbot)	T4 (Info + Chatbot)	p-value joint
Panel A: Narrative groups					
1. Supply-side	0.534	0.018	0.015	-0.002	0.716
		(0.024)	(0.022)	(0.022)	
2. Demand-side	0.128	0.001	0.005	0.010	0.898
		(0.017)	(0.015)	(0.015)	
3. Government mismanagement	0.245	-0.010	-0.044**	-0.019	0.071
		(0.021)	(0.019)	(0.019)	
4. Other / no driver	0.094	-0.010	0.023*	0.010	0.080
		(0.014)	(0.014)	(0.013)	
Panel B: Categories					
5. Tariffs / trade policy	0.150	-0.014	0.020	0.006	0.209
		(0.017)	(0.016)	(0.016)	
6. Government / politics (general)	0.245	-0.010	-0.044**	-0.019	0.071
		(0.021)	(0.019)	(0.019)	
7. Fiscal policy, spending, debt	0.069	0.014	0.000	0.004	0.726
		(0.014)	(0.011)	(0.012)	
8. Monetary policy / interest rates	0.030	0.002	0.008	0.007	0.723
		(0.009)	(0.008)	(0.008)	
9. Supply chain disruptions	0.080	-0.014	-0.024**	-0.011	0.185
		(0.013)	(0.012)	(0.012)	
10. Energy / oil / gas prices	0.085	-0.000	0.014	0.003	0.586
		(0.014)	(0.013)	(0.012)	
11. Wages / labor shortage	0.012	0.018**	0.009*	0.009*	0.055
		(0.007)	(0.006)	(0.006)	
12. Corporate pricing / greed	0.059	0.024*	0.005	0.002	0.251
		(0.013)	(0.011)	(0.011)	
13. Consumer demand	0.029	-0.015*	-0.003	-0.001	0.086
		(0.008)	(0.008)	(0.008)	
14. AI / technology	0.018	0.003	0.004	0.005	0.861
		(0.007)	(0.006)	(0.006)	
15. Geopolitical conflict	0.136	0.006	-0.005	-0.012	0.701
		(0.018)	(0.016)	(0.016)	
16. Climate / weather	0.011	-0.001	-0.005	0.000	0.368
		(0.005)	(0.004)	(0.005)	
17. Housing / rent	0.025	-0.001	0.010	0.002	0.444
		(0.008)	(0.008)	(0.007)	
18. General cost of living (no driver)	0.014	0.001	0.009	-0.000	0.399
		(0.006)	(0.006)	(0.005)	
19. Other / unclear	0.033	-0.014*	-0.002	0.000	0.142
		(0.008)	(0.008)	(0.008)	
20. No substantive response	0.004	0.002	0.003	0.003	0.770
		(0.004)	(0.003)	(0.004)	
Observations			3,855		

Notes: Each row is a linear probability model of an indicator for the LLM-coded posterior inflation-driver narrative (open text) on arm dummies T2–T4 (T1 omitted), the 10 pre-treatment inflation-driver dummies, and demographic controls. Panel A groups the 16 categories of Panel B. The last column reports the p-value of the joint test T2 = T3 = T4 = 0. Excludes blank answers and chatbot-arm respondents who sent no message. Robust standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 9 Treatment Effects on Preferred Government Response

	Control mean	T2 (Information)	T3 (Chatbot)	T4 (Info + Chatbot)	p-value joint
Monetary policy / interest rates	0.135	-0.004	0.021	0.017	0.316
		(0.017)	(0.016)	(0.016)	
Cut spending / control debt	0.128	0.016	-0.014	0.004	0.268
		(0.018)	(0.015)	(0.016)	
Direct household support / subsidies	0.095	0.010	0.037**	0.012	0.062
		(0.016)	(0.014)	(0.014)	
Price controls / antitrust	0.105	0.010	-0.013	-0.017	0.217
		(0.016)	(0.014)	(0.014)	
Lower taxes, tariffs, regulation	0.114	-0.027*	-0.032**	-0.033**	0.093
		(0.015)	(0.014)	(0.014)	
Politics / governance (general)	0.092	0.007	-0.006	-0.005	0.780
		(0.015)	(0.013)	(0.013)	
Supply chain / trade	0.054	-0.010	-0.002	0.022**	0.016
		(0.011)	(0.011)	(0.011)	
Labor, wages, immigration	0.051	-0.005	-0.001	0.016	0.200
		(0.011)	(0.010)	(0.011)	
Energy supply / costs	0.056	-0.015	-0.008	-0.016	0.396
		(0.011)	(0.010)	(0.010)	
Do nothing / market forces	0.010	0.007	0.001	0.007	0.389
		(0.006)	(0.005)	(0.005)	
Regulate AI / technology	0.007	-0.003	0.003	-0.000	0.519
		(0.004)	(0.004)	(0.004)	
Other / unclear	0.116	0.004	-0.005	-0.015	0.570
		(0.017)	(0.015)	(0.015)	
No substantive response	0.039	0.010	0.019*	0.008	0.275
		(0.011)	(0.010)	(0.009)	
Observations			3,857		

Notes: Specification of Table 8, with an indicator for each LLM-coded category of the respondent's preferred government response as the outcome. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

7.2.2 Comprehension and financial literacy

Perhaps conversations with AI (T1 vs T3 and T4) can reverse the spending response observed in the baseline case (T1 vs T2) because AI helps respondents better understand economic/financial concepts (e.g., the difference between real and nominal interest rates). Our estimates (Table 10) suggest that this explanation is not supported by the data. Average treatment effects are close to zero and if anything treatment T4 (Info+Chatbot) appears to worsen responses to financial literacy questions.

7.2.3 Uncertainty

At least since Brainard (1967), uncertainty is often invoked to rationalize a low pass-through from beliefs to actions. In a similar spirit, cognitive uncertainty can attenuate pass-through. To be clear, our focus now is to explain why the pass-through flips from a *significant positive* response when only inflation expectations are changed (T1 vs T2) to a *significant negative*

Table 10 Comprehension, Financial Literacy, and Inaction Threshold: Intention-to-Treat Effects

	Control	T2	T3	T4	p-value	
	mean	(Information)	(Chatbot)	(Info + Chatbot)	T2 = T4	T3 = T4
Panel A: Comprehension						
Real return	0.608	0.038	0.024	0.016	0.316	0.667
		(0.025)	(0.023)	(0.023)		
Purchase timing	0.843	-0.022	-0.015	-0.025	0.887	0.543
		(0.020)	(0.017)	(0.017)		
Real rate	0.414	0.010	-0.014	-0.006	0.489	0.704
		(0.025)	(0.023)	(0.023)		
Combined score (0–3)	1.865	0.026	-0.005	-0.015	0.328	0.776
		(0.047)	(0.042)	(0.043)		
Panel B: Financial literacy						
Compound interest	0.927	-0.013	-0.030**	-0.032**	0.168	0.859
		(0.014)	(0.013)	(0.013)		
Real value	0.845	-0.010	0.002	-0.038**	0.108	0.007
		(0.019)	(0.017)	(0.017)		
Diversification	0.779	0.008	-0.027	-0.035*	0.024	0.620
		(0.021)	(0.019)	(0.019)		
Combined score (0–3)	2.550	-0.015	-0.055	-0.106***	0.011	0.103
		(0.038)	(0.034)	(0.035)		
Panel C: Follow-up survey						
Inflation inaction threshold (pp)	5.959	0.419	-0.316	0.057	0.418	0.300
		(0.502)	(0.429)	(0.431)		

Notes: Each row is a separate regression on arm dummies T2–T4 (T1 omitted) and demographic controls. Panels A and B: linear probability model for a correct answer to each item, and OLS for the combined score (0–3). Panel C: OLS, inflation inaction threshold from the follow-up survey (winsorized at the 1st and 99th percentiles; Figure 10), follow-up respondents only, additionally controlling for days since the main survey. The p-value columns test equality of arm coefficients. Controls: female, college degree, high household income, employed, log baseline non-durable spending, and age-band dummies. Results by chatbot advice appear in Appendix Table A11. Robust standard errors in parentheses. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

response when changes in inflation expectations are coupled with AI guidance (T1 vs T4 below-median). In other words, we are interested in the directional reversal rather than the level of the estimated pass-through. In this context, uncertainty about the inflation outlook or confidence in decisions appears to play a limited role: as we documented in Tables 4 and 5, the treatment effects for both are quite modest.

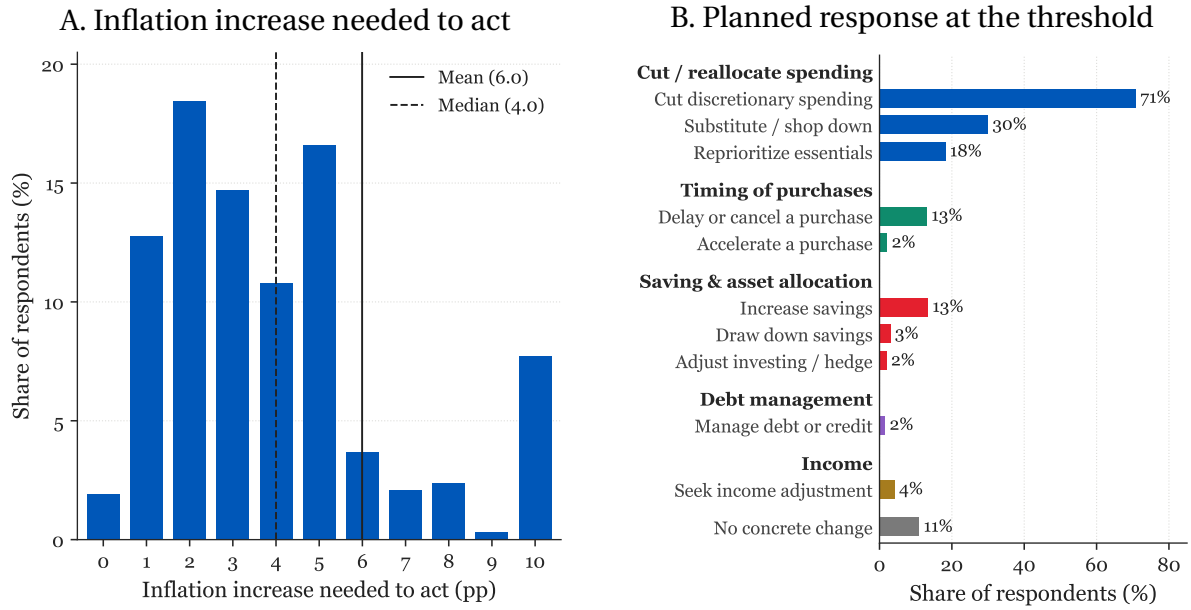
7.2.4 Thresholds

To the extent purchases of durable goods have an important fixed-cost component, the reversal can be driven by a change in the threshold for economic response (e.g., respondents treated with AI can develop a higher threshold for action and therefore are less likely to plan on buying a durable good). To explore this conjecture, the follow-up survey asks how large an increase in inflation would be needed to make respondents change their economic behavior. The median (mean) answer is 4 (6.0) percentage points (Panel A, Figure 10).[17] Once the

[17]Appendix Figure A12 summarizes how the threshold varies with demographics and beliefs. For example, Republicans report a 0.9-percentage-point larger threshold, and so do respondents with more uncertain forecasts.

threshold is crossed, 79% would reduce spending, mostly by cutting discretionary purchases (71%; Panel B). But while thresholds can be an attenuating factor for consumption and thus contribute to low pass-through (indeed, respondents who report a larger threshold took fewer actions in the following month), we find that treatments do not explain variation in the thresholds and hence cannot explain the reversal (Panel C, Table 10).

Figure 10 The Inaction Threshold and the Planned Response at the Threshold



Notes: Follow-up survey, all arms pooled. Panel A: smallest increase in inflation (percentage points) that would make the respondent change their economic behavior, truncated at 10 points for display; vertical lines mark the mean (solid) and median (dashed); $N = 3{,}074$. Panel B: actions at that threshold, coded from open text by GPT-5.6 Luna; categories are not exclusive; $N = 2{,}927$.

7.2.5 Hypothetical inflation shock

One may wonder whether specific advice from AI teaches respondents a more general lesson that can be applied to other settings. To assess this possibility, we ask respondents to describe their reaction to a hypothetical 2pp increase/decrease in inflation. Because this question is asked after treatment, we focus on the within-treatment variation, i.e., estimate the behavioral response by comparing answers in the increase vs. decrease scenarios for a given treatment arm. To this end, we estimate the following specification:

$$\begin{aligned} Response_i = a_0 + b_0 \times I\{i \in Scenario + 2pp\} + \sum_{j=2}^{4} a_j \times I\{i \in Treat\ j\} \\ + \sum_{j=2}^{4} b_j \times I\{i \in Treat\ j\} \times I\{i \in Scenario + 2pp\} + error_i, \end{aligned} \tag{5}$$

where b_2, b_3, b_4 are the coefficients of interest.

We find (Table 11) that higher inflation generally makes consumer spending less likely. For example, moving from -2pp to +2pp scenario for respondents in T2 reduces the probability

Table 11 Hypothetical Inflation Shock: Extensive and Intensive Margins of the Spending Response

	Response category			Direction		% change
	Decrease	Unchanged	Increase	OLS	Ordered probit	OLS
	(1)	(2)	(3)	(4)	(5)	(6)
+2pp shock	0.668***	-0.485***	-0.183***	-0.851***	-1.673***	-13.182***
	(0.028)	(0.033)	(0.025)	(0.041)	(0.098)	(0.787)
T2 (Information)	-0.001	-0.024	0.025	0.026	0.046	-0.097
	(0.019)	(0.035)	(0.032)	(0.040)	(0.072)	(0.643)
T2 × +2pp shock	0.037	-0.016	-0.021	-0.058	-0.137	0.703
	(0.039)	(0.046)	(0.036)	(0.058)	(0.137)	(1.112)
T3 (Chatbot)	0.008	-0.048	0.040	0.033	0.060	-0.051
	(0.018)	(0.032)	(0.029)	(0.037)	(0.066)	(0.581)
T3 × +2pp shock	-0.025	0.052	-0.027	-0.002	0.019	2.149**
	(0.035)	(0.043)	(0.033)	(0.053)	(0.121)	(0.979)
T4 (Info + Chatbot)	0.021	-0.053*	0.032	0.011	0.020	-0.405
	(0.018)	(0.031)	(0.029)	(0.036)	(0.066)	(0.571)
T4 × +2pp shock	-0.019	0.045	-0.026	-0.008	-0.012	1.546
	(0.036)	(0.042)	(0.032)	(0.053)	(0.122)	(0.991)
p-value: T2 × shock = T3 × shock	0.073	0.104	0.848	0.300	0.211	0.139
p-value: T2 × shock = T4 × shock	0.113	0.148	0.872	0.350	0.320	0.395
p-value: T3 × shock = T4 × shock	0.831	0.836	0.970	0.908	0.772	0.472
Control mean, −2pp shock	0.070	0.701	0.228	0.158		1.99
Observations	3,635	3,635	3,635	3,635	3,635	3,858

Notes: Each column regresses the response to a randomized +2 or −2 percentage-point hypothetical inflation shock on a +2pp shock indicator, arm dummies T2–T4 (T1 omitted), and their interactions. Columns (1)–(3): linear probability models for decreasing, leaving unchanged, and increasing the quantity of goods and services purchased. Columns (4)–(5): direction of the change (−1, 0, +1), OLS and ordered probit. Columns (1)–(5) exclude "don't know" responses. Column (6): signed percentage change in purchases, set to 0 for "unchanged" and "don't know" and capped at ±30. Main survey; excludes chatbot-arm respondents who sent no message. Control mean: T1 respondents facing the −2pp shock. Robust standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

of higher spending by 20.4 percentage points (column 3). The corresponding figures for respondents in T3 and T4 are 21.0 and 20.9 percentage points. While responses are broadly similar across the treatment arms (and there is no statistically significant difference in hypothetical responses in the follow-up wave, see Appendix Table A12), our estimates indicate that AI-treated respondents reduce spending by less, which contrasts with the estimated responses of the planned/actual spending documented in Section 7.1. This pattern applies to other hypothetical reactions such as "bring forward purchases," "cut spending, save more," etc. (see Appendix Table A13).

We interpret this evidence as suggesting that respondents follow AI advice but do not necessarily learn from it. This would be consistent with earlier evidence that AI reduces cognitive load and effort (Stadler et al., 2024; Lee et al., 2025) and encourages reliance on its recommendations (Buçinca et al., 2021; Dell'Acqua, 2022), but that potential performance gains from using AI need not translate into learning or improved independent performance (Bastani et al., 2025; Fernandes et al., 2025).

7.2.6 Taking stock

Taken together, the evidence in this section narrows down why a conversation with AI changes what respondents do with their inflation expectations. The conversation does not appear to work by changing how respondents see the economy. It leaves their views on the drivers of inflation and the desired policy response largely unchanged, does not improve their understanding of economic and financial concepts, has only modest effects on uncertainty about the inflation outlook and on confidence in decisions, and does not move the inflation threshold that respondents say would trigger a change in behavior. Nor does the conversation teach respondents a new rule or understanding of the economy. A simple remaining alternative is that respondents report spending plans consistent with the advice they have just received. Realized spending in the follow-up survey, although less precisely estimated, also varies with the direction of the advice, which suggests that the effect is not merely a reporting artifact.

To be clear, these findings do not imply that uncertainty, narratives, comprehension, or inaction thresholds are unimportant for household behavior in general; they may well matter for the level of the pass-through.[18] But because the treatments do not move them, they cannot account for the differences in pass-through across arms.

8 Conclusion

Bertie Wooster, P. G. Wodehouse's amiable hero, rarely solved a problem on his own. When trouble loomed, he turned to his valet, Jeeves, who always knew what to do and who took a dim view of Bertie's extravagant purchases. Many households now have a Jeeves of their own. When they expect higher inflation, they can ask an AI assistant whether to buy a car now or cut back, and, much like Jeeves, it tends to counsel restraint.

We study how such advice shapes the transmission from beliefs to decisions. In a randomized controlled trial, we generate independent variation in inflation expectations and in access to an open-ended conversation with an AI assistant. Respondents generally use the assistant to ask about their own finances rather than the macroeconomy, and the advice they receive is predominantly cautious. Inflation information moves expectations by similar amounts with and without the conversation, but what people do with those expectations differs. For example, without the conversation, higher expected inflation raises planned durable spending. With it, the pass-through falls to roughly zero, and it turns negative where the advice to cut back was strongest. Because the treatments leave uncertainty, narratives, comprehension, and other factors largely unchanged, our evidence points to a distinct

[18] For example, suppose consumers choose between options A and B. They may plan to use A and feel fairly confident about it before a treatment, but they may plan to use B and have the same level of confidence about it after the treatment. Thus, even if the level of confidence (cognitive uncertainty) stays constant, cognitive uncertainty can play a role in determining the level of pass-through.

friction: people may know what will happen without knowing what to do about it.

This matters because standard models and most of their behavioral extensions locate frictions in beliefs: people may be inattentive, misinformed, or uncertain, but given their beliefs, they know how to act. (The notable exception is cognitive uncertainty about decisions, which we measure directly and which our treatments move only modestly.) Our results suggest that many people do not know how to act, and that advice can change what they do without much changing how confident they feel about it. How this friction relates to those already in macroeconomic models is an open question. It may substitute for them (for example, some of the inaction that models attribute to adjustment costs or financial constraints may instead reflect households not knowing what to do), or it may interact with them and amplify them. Distinguishing these possibilities and gauging the quantitative importance of the friction for aggregate dynamics call for theoretical work on how best to model it.

Even before these questions are resolved, the findings have implications for policy. For example, central bank communication focuses on the inflation outlook and the policy path, but the response depends on how households interpret the news. As households turn to AI for interpretations, the design choices of AI providers can shape default actions. Perhaps surprisingly, belt-tightening advice from AI can create a headwind for policymakers who seek to stimulate spending by raising inflation expectations. Managing expectations may thus increasingly require attention to the advice that accompanies them.

Looking further ahead, we can only speculate. For most of history, only the likes of Bertie Wooster could afford a Jeeves. Now anyone with a smartphone can ring for Jeeves and get help living up to Goethe's maxim. Indeed, we may be at the start of a revolution in how households make decisions. Jeeves liked to say that he endeavored to give satisfaction. Whether the new Jeeves does, for households and for the economy as a whole, is a question for the years ahead.

References

Ahmed, Shuaib, Rhosyn Almond, Cameron Belton, Daniel Bogiatzis-Gibbons, Krishane Patel, Manasi Phadnis, Patrick Sholl, and Jackie Spang. 2025. "Money Talks: Lessons from Two LLM Pilots on Consumer Guidance." Financial Conduct Authority FCA Research Note.

Andersen, Steffen, John Y. Campbell, Kasper Meisner Nielsen, and Tarun Ramadorai. 2020. "Sources of Inaction in Household Finance: Evidence from the Danish Mortgage Market." *American Economic Review*, 110(10): 3184–3230.

Andre, Peter, Carlo Pizzinelli, Christopher Roth, and Johannes Wohlfart. 2022. "Subjective Models of the Macroeconomy: Evidence from Experts and Representative Samples." *Review of Economic Studies*, 89(6): 2958–91.

Angeletos, George-Marios, and Chen Lian. 2018. "Forward Guidance without Common Knowledge." *American Economic Review*, 108(9): 2477–2512.

Armantier, Olivier, Scott Nelson, Giorgio Topa, Wilbert van der Klaauw, and Basit Zafar. 2016. "The Price Is Right: Updating Inflation Expectations in a Randomized Price Information Experiment." *Review of Economics and Statistics*, 98(3): 503–523.

Bachmann, Rüdiger, Tim O. Berg, and Eric R. Sims. 2015. "Inflation Expectations and Readiness to Spend: Cross-Sectional Evidence." *American Economic Journal: Economic Policy*, 7(1): 1–35.

Bastani, Hamsa, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakçı, and Rei Mariman. 2025. "Generative AI without Guardrails Can Harm Learning: Evidence from High School Mathematics." *Proceedings of the National Academy of Sciences*, 122(26): e2422633122.

Bick, Alexander, Adam Blandin, and David J. Deming. 2026. "The Rapid Adoption of Generative AI." *Management Science*.

Bilal, Iman Munire, Filippo Giovagnini, Yingcan Carol Wang, Pranav Tewari, Ajan Raj, Yuwei Zhang, and Mei-Chen Zoe Liou. 2026. "From Information to Delegation: Mapping Human-AI Financial Decision Making." arXiv arXiv preprint 2608.02100.

Binder, Carola, and Alex Rodrigue. 2018. "Household Informedness and Long-Run Inflation Expectations: Experimental Evidence." *Southern Economic Journal*, 85(2): 580–598.

Binder, Carola Conces, Rupal Kamdar, and Jane M. Ryngaert. 2024. "Partisan Expectations and COVID-Era Inflation." *Journal of Monetary Economics*, 148: 103649.

Blinder, Alan S., Michael Ehrmann, Jakob de Haan, and David-Jan Jansen. 2024. "Central Bank Communication with the General Public: Promise or False Hope?" *Journal of Economic Literature*, 62(2): 425–57.

Board of Governors. 2026. "Economic Well-Being of U.S. Households in 2025." Board of Governors of the Federal Reserve System Report.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer. 2020. "Overreaction in Macroeconomic Expectations." *American Economic Review*, 110(9): 2748–82.

Brainard, William C. 1967. "Uncertainty and the Effectiveness of Policy." *American Economic Review*, 57(2): 411–425.

Buçinca, Zana, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. "To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making." *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1).

Burke, Mary A., and Ali Ozdagli. 2023. "Household Inflation Expectations and Consumer Spending: Evidence from Panel Data." *Review of Economics and Statistics*, 105(4): 948–961.

Candia, Bernardo. 2026. "Inflation Expectations and Household Spending: Different Patterns in Low- and High-Inflation Settings." Federal Reserve Bank of Cleveland Working Paper 26-26.

Candia, Bernardo, Olivier Coibion, and Yuriy Gorodnichenko. 2020. "Communication and the Beliefs of Economic Agents." NBER Working Paper 27800.

Cavallo, Alberto, Guillermo Cruces, and Ricardo Perez-Truglia. 2017. "Inflation Expectations, Learning, and Supermarket Prices: Evidence from Survey Experiments." *American Economic Journal: Macroeconomics*, 9(3): 1–35.

Celebi, Can, Christine Exley, Soren Harrs, Hannu Kivimaki, Marta Serra-Garcia, and Jeffrey Yusof. 2026. "Mission possible: The collection of high-quality data." *CESifo Working Paper*.

Chatterji, Aaron, Thomas Cunningham, David J. Deming, Zoe Hitzig, Christopher Ong, Carl Yan Shan, and Kevin Wadman. 2025. "How People Use ChatGPT." National Bureau of Economic Research NBER Working Paper 34255.

Choi, Hongseok, Jeongbin Kim, Matthew Kovach, Kyu-Min Lee, Euncheol Shin, and Hector Tzavellas. 2026. "Do People Follow AI Advice? Evidence from a Pension Portfolio Choice Experiment." arXiv arXiv preprint 2608.11371.

Chopra, Felix, and Ingar Haaland. 2023. "Conducting Qualitative Interviews with AI." CESifo Working Paper 10666.

Chopra, Felix, Ingar Haaland, Nicolas Roever, and Christopher Roth. 2026. "Evaluating Behavioral Interventions at Scale with AI." Working paper.

Choukhmane, Taha, Tim de Silva, Weidong Lin, and Matthew Akuzawa. 2026. "AI Financial Advice: Supply, Demand, and Life Cycle Implications." National Bureau of Economic Research NBER Working Paper 35574.

Coibion, Olivier, and Yuriy Gorodnichenko. 2012. "What Can Survey Forecasts Tell Us about Information Rigidities?" *Journal of Political Economy*, 120(1): 116–59.

Coibion, Olivier, and Yuriy Gorodnichenko. 2015. "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts." *American Economic Review*, 105(8): 2644–78.

Coibion, Olivier, and Yuriy Gorodnichenko. 2026. *Expectations Matter: The New Causal Macroeconomics of Surveys and Experiments*. Princeton, NJ:Princeton University Press.

Coibion, Olivier, Dimitris Georgarakos, Yuriy Gorodnichenko, and Maarten van Rooij. 2023. "How Does Consumption Respond to News about Inflation? Field Evidence from a Randomized Control Trial." *American Economic Journal: Macroeconomics*, 15(3): 109–152.

Coibion, Olivier, Dimitris Georgarakos, Yuriy Gorodnichenko, Geoff Kenny, and Michael Weber. 2024. "The Effect of Macroeconomic Uncertainty on Household Spending." *American Economic Review*, 114(3): 645–677.

Coibion, Olivier, Yuriy Gorodnichenko, and Michael Weber. 2022. "Monetary Policy Communications and Their Effects on Household Inflation Expectations." *Journal of Political Economy*, 130(6): 1537–1584.

Coibion, Olivier, Yuriy Gorodnichenko, and Tiziano Ropele. 2020. "Inflation Expectations and Firm Decisions: New Causal Evidence." *Quarterly Journal of Economics*, 135(1): 165–219.

Crump, Richard K., Stefano Eusepi, Andrea Tambalotti, and Giorgio Topa. 2022. "Subjective Intertemporal Substitution." *Journal of Monetary Economics*, 126: 118–133.

D'Acunto, Francesco, Daniel Hoang, Maritta Paloviita, and Michael Weber. 2023. "IQ, Expectations, and Choice." *Review of Economic Studies*, 90(5): 2292–2325.

D'Acunto, Francesco, Nagpurnanand Prabhala, and Alberto G. Rossi. 2019. "The Promises and Pitfalls of Robo-Advising." *Review of Financial Studies*, 32(5): 1983–2020.

D'Acunto, Francesco, Ulrike Malmendier, and Michael Weber. 2021. "Gender Roles Produce Divergent Economic Expectations." *Proceedings of the National Academy of Sciences*, 118(21): e2008534118.

Dell'Acqua, Fabrizio. 2022. "Falling Asleep at the Wheel: Human/AI Collaboration in a Field Experiment on HR Recruiters." Laboratory for Innovation Science at Harvard, Harvard Business School Working Paper.

Enke, Benjamin, and Thomas Graeber. 2023. "Cognitive Uncertainty." *Quarterly Journal of Economics*, 138(4): 2021–67.

Enke, Benjamin, Thomas Graeber, Ryan Oprea, and Jeffrey Yang. forthcoming. "Behavioral Attenuation." *Journal of Political Economy*.

Fernandes, Daniela, Steeven Villa, Salla Nicholls, Otso Haavisto, Daniel Buschek, Albrecht Schmidt, Thomas Kosch, Chenxinran Shen, and Robin Welsch. 2025. "AI Makes You

Smarter but None the Wiser: The Disconnect between Performance and Metacognition." *Computers in Human Behavior*, 108779.

Financial Conduct Authority. 2026. "Young Investors Trust AI More Than TV or Celebrities." *Press release.*

FINRA Investor Education Foundation. 2024. "The Machines Are Coming (with Personal Finance Information). Do We Trust Them?" FINRA Investor Education Foundation Report.

Gabaix, Xavier. 2020. "A Behavioral New Keynesian Model." *American Economic Review*, 110(8): 2271–2327.

Gallup. 2026. "Gallup–Edward Jones Survey on Sources of Financial Advice." *Reported in Associated Press, "Some US Adults Are Using AI for Financial Guidance but Few Trust It, Gallup Poll Finds," August 7, 2026.*

Geiecke, Friedrich, and Xavier Jaravel. forthcoming. "Conversations at Scale: Robust AI-led Interviews." *Review of Economic Studies.*

Georgarakos, Dimitris, Yuriy Gorodnichenko, Olivier Coibion, and Geoff Kenny. 2024. "The Causal Effects of Inflation Uncertainty on Households' Beliefs and Actions." National Bureau of Economic Research, Inc NBER Working Paper 33014.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus. 2021. "Five Facts about Beliefs and Portfolios." *American Economic Review*, 111(5): 1481–1522.

Greubel, Joshua, Henrik Guhling, and Fabian Herweg. 2026. "AI Persuasion and Financial-Decision Making: Experimental Evidence on Dominated Investment Choices." CESifo CESifo Working Paper 12925.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart. 2023. "Designing Information Provision Experiments." *Journal of Economic Literature*, 61(1): 3–40.

Horton, John J. 2023. "Large Language Models as Simulated Economic Agents: What Can We Learn from Homo Silicus?" *NBER Working Paper*, , (31122).

Huang, Jin, Diego Ferreras Garrucho, Yutong Xie, Walter M Yuan, Qiaozhu Mei, Chen Lian, and Jonathon Hazell. 2026. "HouseholdBench: Evaluating Large Language Models as Predictors of Household Economic Behavior." *NBER Working Paper.*

Humlum, Anders, and Emilie Vestergaard. 2025. "The Unequal Adoption of ChatGPT Exacerbates Existing Inequalities among Workers." *Proceedings of the National Academy of Sciences*, 122(1): e2414972121.

Kamdar, Rupal, and Walker Ray. 2024. "Polarized Expectations, Polarized Consumption." Financial Markets Group, London School of Economics FMG Discussion Paper 902.

Korenok, Oleg, David Munro, and Jiayi Chen. 2026. "Inflation and Attention Thresholds." *Review of Economics and Statistics,* 108(3): 842–850.

Kostyshyna, Olena, and Luba Petersen. 2024. "The Effect of Inflation Uncertainty on Household Expectations and Spending." National Bureau of Economic Research NBER Working Paper 32939.

Lee, Hao-Ping (Hank), Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson. 2025. "The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers." *CHI '25*. Yokohama, Japan.

Lusardi, Annamaria, and Olivia S. Mitchell. 2008. "Planning and Financial Literacy: How Do Women Fare?" *American Economic Review*, 98(2): 413–417.

Mullainathan, Sendhil, Markus Noeth, and Antoinette Schoar. 2012. "The Market for Financial Advice: An Audit Study." NBER Working Paper 17929.

Otis, Nicholas G., Solène Delecourt, Katelyn Cranney, and Rembrand Koning. 2024. "Global Evidence on Gender Gaps and Generative AI." Harvard Business School Working Paper 25-023.

Pew Research Center. 2026. "Americans and AI 2026: Chatbots, Smart Devices and Views on Impact." *Report, June 17, 2026*.

Pfäuti, Oliver. forthcoming. "The Inflation Attention Threshold and Inflation Surges." *American Economic Review*.

Schnorpfeil, Philip, Michael Weber, and Andreas Hackethal. 2025. "Inflation and Trading." *Journal of Financial Economics,* 173: 104166.

Shafir, Eldar, Peter Diamond, and Amos Tversky. 1997. "Money Illusion." *Quarterly Journal of Economics,* 112(2): 341–74.

Stadler, Matthias, Maria Bannert, and Michael Sailer. 2024. "Cognitive Ease at a Cost: LLMs Reduce Mental Effort but Compromise Depth in Student Scientific Inquiry." *Computers in Human Behavior,* 160: 108386.

Stantcheva, Stefanie. 2024. "Why Do We Dislike Inflation?" *Brookings Papers on Economic Activity,* , (Spring): 1–46.

Stokey, Nancy L. 2009. *The Economics of Inaction: Stochastic Control Models with Fixed Costs.* Princeton, NJ:Princeton University Press.

Weber, Michael, Bernardo Candia, Hassan Afrouzi, Tiziano Ropele, Rodrigo Lluberas, Serafin Frache, Brent H. Meyer, Saten Kumar, Yuriy Gorodnichenko, Dimitris Georgarakos, Olivier Coibion, Geoff Kenny, and Jorge Ponce. 2025. "Tell Me Something I Don't Already Know: Learning in Low- and High-Inflation Settings." *Econometrica*, 93(1): 229–264.

Westwood, Sean J. 2025. "The Potential Existential Threat of Large Language Models to Online Survey Research." *Proceedings of the National Academy of Sciences*, 122(47): e2518075122.

Online Appendix

A Additional Results

A.1 Additional Tables

Table A1 Demographic Composition of the Analysis Sample

	Sample (%)	U.S. population (%)
Gender		
Male	46.3	49.0
Female	52.6	51.0
Employment status		
Employed full-time	52.4	42.9
Employed part-time	14.7	11.9
Self-employed	13.6	6.2
Unemployed and looking for work	8.2	2.9
Not in the labor force (student, retired, homemaker, etc.)	10.8	36.0
Educational attainment		
Less than high school	0.4	10.3
High school diploma or equivalent	12.3	27.1
Some college, no degree	18.3	20.0
Associate's degree	10.2	8.5
Bachelor's degree	39.5	21.0
Master's degree	15.4	9.4
Doctoral degree (PhD, MD, JD, etc.)	3.7	3.7
Household income		
Less than $25,000	8.7	14.4
$25,000 – $49,999	18.6	16.2
$50,000 – $74,999	21.8	15.6
$75,000 – $99,999	17.8	12.6
$100,000 – $149,999	18.6	17.8
$150,000 – $199,999	7.5	9.7
$200,000 or more	4.8	13.7
Age		
18-24	8.5	11.7
25-34	29.5	17.3
35-44	29.9	17.2
45-54	18.2	15.3
55-64	9.9	15.6
65 or above	4.1	22.9
Political affiliation		
Democrat	44.5	27.0
Republican	23.5	27.0
Independent	28.4	45.0
Other	2.5	–
Prefer not to answer	1.0	–
Observations	3,870	

Notes: Share of respondents in each category. U.S. population shares: ACS 2024 and Gallup 2025. $N = 3{,}870$.

Table A2 Covariate Balance Across Arms

	Control (T1)	Info (T2)	Chatbot (T3)	Info + Chatbot (T4)	Sig. pairs (of 6)
Female	0.52	0.53	0.52	0.53	0
Age 65 or above	0.04	0.03	0.04	0.04	0
College degree	0.60	0.57	0.61	0.56	1
Income ≥ $100k	0.30	0.30	0.32	0.31	0
Employed	0.80	0.80	0.81	0.81	0
Prior inflation forecast (%)	4.11	4.04	4.00	4.02	0
Prior forecast uncertainty	2.54	2.58	2.53	2.58	0
Prior decision confidence	54.26	54.22	55.34	53.65	0
AI use frequency (0–5)	2.17	2.16	2.24	2.19	0
Trust in AI advice (0–4)	2.02	1.97	2.02	1.98	0
Perceived AI accuracy vs. other sources (−2 to 2)	0.28	0.23	0.21	0.19	1
AI use changed respondent's view (0–10)	4.61	4.83	4.65	4.58	1
Macro info-source consultation frequency (0–6)	2.66	2.69	2.70	2.68	0
Preference rank of AI (1–5)	3.18	3.19	3.16	3.23	0
WTP for AI advice ($)	11.70	11.59	10.95	11.28	0
WTP for expert advice ($)	48.14	42.39	46.36	48.81	0
WTP for Federal Reserve advice ($)	20.05	17.85	18.62	19.91	0
Observations	727	730	1,207	1,206	

Significant pairwise imbalances at 5%: 3/102 (2.9%)
Joint F-test, Information indicator on all covariates: $F = 0.60$, $p = 0.896$
Joint F-test, Chatbot indicator on all covariates: $F = 0.90$, $p = 0.579$
Joint F-test (exploratory, not PAP-registered), Info × Chatbot indicator on all covariates: $F = 0.59$, $p = 0.904$

Notes: Arm means of each pre-randomization covariate and the number of pairwise arm comparisons (of six) significant at the 5% level, two-sided t-test. Joint F-tests regress the treatment indicator on the full covariate vector. $N = 3{,}870$.

Table A3 Inflation Expectations: Sample Moments and U.S. Household-Survey Benchmarks

	Mean	SD	p10	p25	p50	p75	p90	N
Panel A: All respondents, pre-treatment								
Inflation forecast, implied mean (%)	4.03	2.49	1.40	2.60	3.55	5.25	7.30	3,870
Inflation forecast, implied uncertainty (pp)	2.56	1.66	0.94	1.47	2.24	3.20	4.52	3,870
Decision confidence (0–100)	54.40	25.23	20.00	36.00	55.00	73.00	87.00	3,870
Panel B: Control group (T1), post-treatment								
Inflation forecast, point prediction (%)	5.26	6.33	2.00	3.00	4.00	5.00	8.00	727
Inflation forecast, scenario-implied uncertainty (pp)	1.70	2.10	0.41	0.70	1.18	1.79	2.94	727

Panel C: Comparison with U.S. household surveys, July 2026

	This survey	SCE	Michigan
Density mean, median (%)	3.55	3.63	–
25th percentile	2.60	2.19	–
75th percentile	5.25	6.00	–
Uncertainty (interquartile range), median	2.89	2.71	–
Mean probability (%) that inflation is …			
below 0%	7.9	16.0	–
0–2%	16.1	11.4	–
2–4%	37.7	26.2	–
above 4%	38.4	46.5	–
Point forecast, median (%)	4.00	4.11	4.20

Notes: Panel A: all respondents, before randomization; the implied mean and uncertainty are the mean and standard deviation of each respondent's density forecast. Panel B: control group (T1), post-treatment; the point forecast is directly elicited and the uncertainty is implied by the three-scenario forecast. Panels A–B are winsorized at the 1st and 99th percentiles. Panel C: main wave against the July 2026 New York Fed Survey of Consumer Expectations (SCE) and the University of Michigan Surveys of Consumers; density rows use the raw density forecast of all respondents, and the point forecast is the control group's. The SCE fits a generalized beta distribution to each density, whereas our mean values the bins at their midpoints (open tails at ±12). Interquartile ranges: the SCE uses the fitted density; ours assumes probability is uniform within each bin, with open tails closed at ±16.

Table A4 Pre-Treatment Beliefs and Demographics

	(1) Implied mean	(2) Implied uncertainty	(3) Decision confidence
Female	0.311***	0.015	-6.883***
	(0.079)	(0.053)	(0.793)
College degree	-0.136	0.049	1.945**
	(0.086)	(0.056)	(0.860)
High income	-0.150*	-0.120**	3.048***
	(0.084)	(0.058)	(0.866)
Employed	-0.295***	0.098	5.731***
	(0.110)	(0.066)	(1.086)
Age	0.022	-0.032**	-0.498**
	(0.018)	(0.012)	(0.196)
Age^2/100	-0.033	0.006	0.665***
	(0.020)	(0.014)	(0.227)
Democrat	0.117	0.039	-1.068
	(0.094)	(0.058)	(0.936)
Republican	-0.927***	0.253***	7.121***
	(0.108)	(0.076)	(1.058)
Observations	3,870	3,870	3,870
Mean of outcome	4.033	2.558	54.397
R^2	0.043	0.044	0.060

Notes: OLS, robust standard errors in parentheses. Outcomes, all elicited before randomization: density-implied prior 1-year inflation mean (col. 1) and uncertainty (col. 2), both winsorized 1/99, and prior decision confidence (0–100, col. 3). Omitted party category: Independent, other, no answer. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table A5 Determinants of Prior Decision Confidence

	(1)	(2)	(3)	(4)
Female	-6.99***	-6.68***	-6.28***	-6.21***
	(0.80)	(0.79)	(0.80)	(0.78)
College degree	1.90**	1.78**	1.43*	1.35
	(0.86)	(0.85)	(0.86)	(0.85)
High household income	2.88***	2.65***	2.61***	1.94**
	(0.86)	(0.86)	(0.85)	(0.83)
Employed	5.16***	4.98***	4.97***	3.50***
	(1.08)	(1.07)	(1.07)	(1.07)
Age 45+	2.06**	1.50*	1.51*	0.84
	(0.88)	(0.89)	(0.89)	(0.88)
Democrat	-1.01	-0.87	-0.63	-0.46
	(0.94)	(0.93)	(0.93)	(0.91)
Republican	7.22***	6.48***	6.05***	4.32***
	(1.06)	(1.06)	(1.06)	(1.06)
Prior inflation forecast (%)		-0.95***	-0.87***	-0.79***
		(0.18)	(0.18)	(0.17)
Prior forecast uncertainty (pp)		-0.66***	-0.65***	-1.10***
		(0.24)	(0.24)	(0.24)
Narrative: government mismanagement			-4.35***	-3.72***
			(0.89)	(0.87)
Narrative: any supply-side cause			-1.25	-2.31**
			(1.18)	(1.16)
Narrative: any demand-side cause			-0.49	-1.21
			(1.07)	(1.05)
Prior AI usage frequency (0–5)				1.42***
				(0.29)
Trust in AI advice (0–4)				4.20***
				(0.60)
Observations	3,870	3,870	3,870	3,870
Mean of outcome	54.40	54.40	54.40	54.40
R^2	0.058	0.067	0.073	0.109

Notes: OLS, robust (HC1) standard errors in parentheses. The dependent variable is prior decision confidence (0–100). Column (1) includes demographics and partisanship (omitted: Independent, other, no answer). Column (2) adds the density-implied prior forecast and forecast uncertainty, both winsorized 1/99. Column (3) adds indicators for the pre-treatment inflation-driver question. Column (4) adds prior AI usage frequency and trust in AI advice. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table A6 2SLS: Posterior Inflation Forecast on the Allocation of a Hypothetical Windfall

	(1) Savings/current acct.	(2) Stocks	(3) Mutual funds	(4) Retirement/pension	(5) Bonds	(6) Gold/metals	(7) Crypto
Posterior inflation forecast	3.2**	-2.1**	-1.5*	0.1	-0.1	0.6	-0.1
	(1.5)	(0.9)	(0.8)	(0.8)	(0.3)	(0.5)	(0.3)
First-stage F (robust)	77.4	77.4	77.4	77.4	77.4	77.4	77.4
Observations	3,430	3,430	3,430	3,430	3,430	3,430	3,430

Notes: Specification, sample, and jackknife trim of Table 6, with each category's share (0–100) of the hypothetical $10,000 windfall as the outcome. The risky share of Table 6 is the sum of columns 2, 3, and 7. Standard errors in parentheses. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table A7 Persistence of AI Trust, Accuracy, and Usage to the Follow-Up Survey

	(1) AI trust	(2) AI trust	(3) AI accuracy	(4) AI use freq. (wk)	(5) View changed
Main-survey AI trust	0.751***	0.751***			
	(0.014)	(0.014)			
Main-survey AI accuracy			0.582***		
			(0.018)		
T2 (Information)		0.011	-0.002	-0.105*	-0.222
		(0.036)	(0.044)	(0.060)	(0.220)
T3 (Chatbot)		0.084***	0.099**	0.008	-0.263
		(0.032)	(0.039)	(0.056)	(0.203)
T4 (Info + Chatbot)		0.056*	0.066*	0.015	-0.225
		(0.033)	(0.040)	(0.056)	(0.199)
Arm dummies	No	Yes	Yes	Yes	Yes
Control group mean	1.813	1.813	0.073	0.965	3.538
p-value: T2 = T4		0.162	0.075	0.024	0.988
p-value: T3 = T4		0.288	0.318	0.870	0.825
Observations	3,074	3,074	3,049	3,054	1,733
R^2	0.543	0.544	0.349	0.042	0.040

Notes: OLS on the arm dummies T2–T4 (T1 omitted), demographic controls, and days between the two surveys; the p-value rows test equality of arm coefficients. Columns 1–2: follow-up AI trust (0–4, six situations) on its main-survey value. Column 3: follow-up perceived AI accuracy on its main-survey value; the follow-up compares AI to four named sources, the main survey to four generic ones. Column 4: follow-up weekly AI-use frequency. Column 5: extent AI use changed a financial decision in the past week (0–10), among respondents who used AI that week. Robust standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A8 Follow-Up Chatbot Novelty: Engagement by Original Main-Survey Arm

	(1) Helpfulness (1-10)	(2) User turns
Information	-0.074	-0.140*
	(0.140)	(0.082)
Chatbot	0.084	-0.039
	(0.128)	(0.073)
Information × Chatbot	0.093	0.087
	(0.177)	(0.103)
Observations	3,074	3,074
Control group mean (original arm)	6.329	3.355
R^2	0.015	0.013

Notes: Follow-up chatbot helpfulness rating (1–10) and number of user turns, regressed on the original main-survey arm (Chatbot = 1 for T3 and T4). Follow-up respondents with a non-missing transcript. Robust standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A9 Follow-Up Chatbot Helpfulness on AI Usage and Main-Survey Rating

	(1) All arms	(2) Chatbot arms only
AI usage frequency, follow-up survey (0–3)	0.631***	0.257***
	(0.039)	(0.044)
Chatbot helpfulness rating, main survey (1–10)		0.555***
		(0.022)
Control group mean (Control arm)	6.35	–
Chatbot-only arm mean	–	6.43
Observations	3,023	1,881
R^2	0.080	0.364

Notes: OLS. Outcome: the follow-up chatbot's helpfulness rating (1–10). Column 1: follow-up respondents matched to a main-survey record. Column 2: adds the main-survey helpfulness rating, T3 and T4 only. AI usage is the follow-up past-week frequency item (0–3). No demographic controls. The sample does not apply the exclusions of Section 2.1. Robust (HC1) standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A10 Planned vs. Realized Spending: Pairwise Correlations

Planned vs. realized	Pearson r	p-value	N
Durable spending (log)	0.302	0.000	3,074
Non-durable spending (log)	0.319	0.000	3,074
Durable goods (category, -2..+2)	0.235	0.000	3,074
Food at home (category, -2..+2)	0.245	0.000	3,074
Restaurants (category, -2..+2)	0.378	0.000	3,074
Non-durable goods (category, -2..+2)	0.257	0.000	3,074
Leisure (category, -2..+2)	0.341	0.000	3,074

Notes: Pearson correlations between main-survey planned or expected measures and their realized follow-up counterparts, follow-up respondents. First two rows: stated dollar plan and realized spending (both in logs). Remaining rows: expected and realized consumption change by category (−2 to +2).

Table A11 Comprehension, Financial Literacy, and Inaction Threshold, by Chatbot Spending Advice

	T3 (Chatbot)		T4 (Info + Chatbot)		p-value: below = above	
	Below	Above	Below	Above	T3	T4
Panel A: Comprehension						
Real return	0.005	0.050*	0.010	0.022	0.105	0.683
	(0.026)	(0.027)	(0.026)	(0.028)		
Purchase timing	-0.031	0.002	-0.034	-0.014	0.129	0.372
	(0.020)	(0.020)	(0.020)	(0.021)		
Real rate	-0.022	-0.003	-0.018	0.008	0.500	0.357
	(0.026)	(0.027)	(0.026)	(0.028)		
Combined score (0–3)	-0.049	0.048	-0.041	0.017	0.058	0.279
	(0.049)	(0.050)	(0.049)	(0.052)		
Panel B: Financial literacy						
Compound interest	-0.038**	-0.021	-0.053***	-0.005	0.315	0.006
	(0.016)	(0.015)	(0.016)	(0.015)		
Real value	0.005	0.001	-0.035*	-0.042**	0.877	0.764
	(0.019)	(0.020)	(0.020)	(0.021)		
Diversification	-0.042*	-0.006	-0.044**	-0.025	0.131	0.454
	(0.022)	(0.022)	(0.022)	(0.023)		
Combined score (0–3)	-0.076*	-0.025	-0.132***	-0.072*	0.247	0.198
	(0.040)	(0.041)	(0.041)	(0.043)		
Panel C: Follow-up survey						
Inflation inaction threshold (pp)	-0.377	-0.229	0.067	0.114	0.774	0.928
	(0.494)	(0.509)	(0.492)	(0.519)		

Notes: Specification of Table 10, with T3 and T4 each split at a GPT-6.1 Sol advice score of −2, as in Table 7: "below" is a score of −2 or lower, "above" a score above −2. The p-value columns test that the two coefficients are equal within T3 and within T4. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A12 Extensive and Intensive Margin Response to Hypothetical Inflation Shock: Follow-Up

	Response category			Direction		% change
	Decrease	Unchanged	Increase	OLS	Ordered probit	OLS
	(1)	(2)	(3)	(4)	(5)	(6)
+2pp shock	0.711***	-0.446***	-0.265***	-0.976***	-1.913***	-15.281***
	(0.030)	(0.037)	(0.030)	(0.048)	(0.119)	(0.971)
T2 (Information)	0.011	0.053	-0.063*	-0.074	-0.130	-0.410
	(0.024)	(0.040)	(0.037)	(0.048)	(0.084)	(0.837)
T2 × +2pp shock	-0.018	-0.033	0.052	0.070	0.122	1.111
	(0.043)	(0.052)	(0.041)	(0.066)	(0.158)	(1.361)
T3 (Chatbot)	0.007	0.005	-0.011	-0.018	-0.032	-0.091
	(0.022)	(0.037)	(0.035)	(0.045)	(0.078)	(0.758)
T3 × +2pp shock	0.011	-0.016	0.005	-0.007	-0.044	0.953
	(0.037)	(0.046)	(0.038)	(0.060)	(0.144)	(1.218)
T4 (Info + Chatbot)	0.010	0.001	-0.011	-0.021	-0.037	0.371
	(0.022)	(0.037)	(0.035)	(0.045)	(0.079)	(0.750)
T4 × +2pp shock	-0.036	0.021	0.015	0.052	0.126	0.512
	(0.038)	(0.047)	(0.039)	(0.061)	(0.142)	(1.220)
p-value: T2 × shock = T3 × shock	0.439	0.712	0.188	0.187	0.242	0.896
p-value: T2 × shock = T4 × shock	0.647	0.246	0.314	0.757	0.977	0.620
p-value: T3 × shock = T4 × shock	0.150	0.357	0.751	0.266	0.170	0.672
Control mean, −2pp shock	0.086	0.604	0.309	0.223		2.12
Observations	2,937	2,937	2,937	2,937	2,937	3,065

Notes: Specification of Table 11 in the follow-up survey, with the shock direction re-drawn independently of the main survey.

Table A13 Hypothetical Inflation Shock: Behavioral Responses by Item

Panel A: Spending and shopping

	Bring fwd. durables (1)	Cut spend., save more (2)	Shop around (3)	Cheaper substitutes (4)	Switch stores (5)	Shop online (6)	Save less, liquidate (7)	More credit (8)
+2pp shock	-0.494***	0.380***	0.868***	1.084***	0.658***	-0.332**	0.388***	0.238*
	(0.124)	(0.110)	(0.092)	(0.104)	(0.111)	(0.131)	(0.128)	(0.135)
T2 (Information)	0.013	-0.146	-0.008	-0.097	-0.004	0.139	-0.073	0.081
	(0.105)	(0.106)	(0.094)	(0.101)	(0.100)	(0.108)	(0.115)	(0.118)
T2 × +2pp shock	-0.288	0.333**	-0.091	0.044	-0.027	-0.152	0.135	0.012
	(0.176)	(0.157)	(0.140)	(0.148)	(0.161)	(0.169)	(0.178)	(0.188)
T3 (Chatbot)	-0.032	-0.151*	-0.084	-0.078	0.012	-0.064	-0.041	0.077
	(0.090)	(0.091)	(0.088)	(0.095)	(0.094)	(0.106)	(0.104)	(0.109)
T3 × +2pp shock	-0.260*	0.209	-0.052	0.008	-0.116	0.030	0.085	-0.005
	(0.149)	(0.140)	(0.126)	(0.133)	(0.147)	(0.160)	(0.161)	(0.173)
T4 (Info + Chatbot)	-0.057	-0.020	-0.097	-0.027	0.006	0.013	-0.094	-0.014
	(0.095)	(0.094)	(0.084)	(0.098)	(0.099)	(0.102)	(0.104)	(0.111)
T4 × +2pp shock	-0.075	0.038	-0.051	0.001	-0.077	-0.166	0.271*	-0.051
	(0.158)	(0.144)	(0.124)	(0.137)	(0.150)	(0.157)	(0.160)	(0.170)
p-value: T2 × shock = T3 × shock	0.853	0.378	0.777	0.785	0.561	0.192	0.754	0.923
p-value: T2 × shock = T4 × shock	0.180	0.042	0.770	0.757	0.747	0.917	0.384	0.707
p-value: T3 × shock = T4 × shock	0.150	0.178	0.994	0.960	0.783	0.115	0.175	0.758
p-value: all arm × shock = 0	0.181	0.092	0.932	0.989	0.866	0.344	0.335	0.980
Control mean, −2pp shock	0.094	0.308	0.298	0.014	-0.327	-0.017	-0.374	-0.604
Observations	1,548	1,498	1,491	1,518	1,458	1,450	1,490	1,461

Panel B: Saving, portfolio and labour income

	Gold, real estate (9)	More cash savings (10)	Less risky portfolio (11)	Ask raise, new job (12)	Other extra income (13)	Mean of seen items (14)
+2pp shock	-0.325**	-0.261**	0.322***	0.398***	0.732***	0.284***
	(0.131)	(0.118)	(0.122)	(0.124)	(0.122)	(0.045)
T2 (Information)	0.086	-0.136	0.050	-0.082	-0.025	-0.018
	(0.116)	(0.109)	(0.106)	(0.105)	(0.123)	(0.041)
T2 × +2pp shock	-0.107	0.141	-0.172	0.394**	0.261	0.044
	(0.184)	(0.166)	(0.165)	(0.173)	(0.170)	(0.063)
T3 (Chatbot)	0.012	-0.011	0.000	0.044	-0.055	-0.027
	(0.109)	(0.107)	(0.099)	(0.096)	(0.111)	(0.039)
T3 × +2pp shock	0.021	0.076	-0.027	-0.070	0.205	0.015
	(0.169)	(0.152)	(0.150)	(0.157)	(0.151)	(0.057)
T4 (Info + Chatbot)	0.096	0.071	-0.062	-0.000	0.098	-0.008
	(0.105)	(0.102)	(0.104)	(0.100)	(0.109)	(0.040)
T4 × +2pp shock	-0.264	0.013	0.039	0.131	0.042	-0.022
	(0.165)	(0.152)	(0.156)	(0.160)	(0.151)	(0.058)
p-value: T2 × shock = T3 × shock	0.449	0.669	0.306	0.003	0.707	0.601
p-value: T2 × shock = T4 × shock	0.341	0.397	0.153	0.096	0.137	0.235
p-value: T3 × shock = T4 × shock	0.054	0.638	0.612	0.150	0.193	0.454
p-value: all arm × shock = 0	0.216	0.801	0.535	0.022	0.254	0.683
Control mean, −2pp shock	-0.431	0.313	-0.137	-0.150	0.064	-0.070
Observations	1,448	1,487	1,499	1,473	1,469	3,858

Notes: Main survey. After the hypothetical ±2pp inflation shock, respondents rated how likely they are to adopt each behavior relative to their current plan, from "much less likely" (−2) to "much more likely" (+2). Each respondent saw a random 5 of the 13 items. Column (14): the respondent's mean over the items shown. Regressors as in Table 11. Excludes chatbot-arm respondents who sent no message. Control mean: T1 respondents facing the −2pp shock. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A14 Belief Pass-Through into Spending by Chatbot Advice: Split at the Median Advice Score per Action

	Log planned spending		Risky	Log realized spending	
	Durable	Nondurable	share	Durable	Nondurable
	(1)	(2)	(3)	(4)	(5)
Panel A: T1 and T2 (Information)					
Posterior inflation forecast	0.212**	0.116*	0.3	-0.118	-0.103
	(0.098)	(0.068)	(1.5)	(0.101)	(0.064)
First-stage *F*	141.3	150.3	109.0	121.5	124.5
Observations	1,315	1,321	1,218	1,040	1,051
Panel B: T1 and T3 (Chatbot)					
Posterior inflation forecast	0.018	-0.618**	-8.2	0.779***	-0.079
	(0.375)	(0.301)	(7.3)	(0.270)	(0.156)
First-stage *F*	6.7	5.9	3.1	14.6	12.2
Observations	1,733	1,731	1,656	1,371	1,366
Panel C: T1 and T4 (Information + Chatbot)					
Posterior inflation forecast	-0.023	0.083	-0.9	-0.065	-0.063
	(0.121)	(0.083)	(1.7)	(0.126)	(0.071)
First-stage *F*	97.0	93.5	83.3	81.1	79.9
Observations	1,733	1,736	1,647	1,377	1,375
Panel D: T1 and T4, below-median advice score per action					
Posterior inflation forecast	-0.106	0.156*	-2.5	-0.096	-0.124*
	(0.118)	(0.084)	(1.9)	(0.113)	(0.070)
First-stage *F*	88.9	89.0	63.5	74.5	73.9
Observations	1,150	1,151	1,056	901	904
Panel E: T1 and T4, above-median advice score per action					
Posterior inflation forecast	0.259	0.189	-4.1	0.240	0.004
	(0.167)	(0.119)	(2.8)	(0.183)	(0.090)
First-stage *F*	41.8	42.3	31.6	35.9	42.4
Observations	1,167	1,173	1,090	917	923
Panel F: T1, T3, and T4, below-median advice score per action					
Posterior inflation forecast	-0.034	0.128	-3.5*	0.047	-0.112*
	(0.114)	(0.081)	(1.9)	(0.111)	(0.068)
First-stage *F*	49.3	48.7	33.7	40.6	42.1
Observations	1,679	1,678	1,588	1,326	1,331
Panel G: T1, T3, and T4, above-median advice score per action					
Posterior inflation forecast	0.355**	0.037	-9.8***	0.623***	0.160*
	(0.157)	(0.107)	(2.8)	(0.174)	(0.086)
First-stage *F*	32.6	32.1	24.0	29.8	31.5
Observations	1,689	1,689	1,616	1,333	1,329
Panel H: Equality of posterior-forecast coefficients (Wald p-values)					
A = C	0.038**	0.674	0.431	0.648	0.567
D = E	0.019**	0.772	0.558	0.037**	0.130
A = D	0.006***	0.616	0.138	0.846	0.768
A = F	0.046**	0.883	0.060*	0.176	0.899
A = G	0.377	0.469	0.001***	0.000***	0.005***
F = G	0.021**	0.437	0.043**	0.001***	0.003***
Pre-treatment spending control	Yes	Yes	No	Yes	Yes

Notes: Specification, samples, and Panels A–C as in Table 7. Panels D–G split T3 and T4 respondents at the pooled median (−0.375) of the GPT-6.1 Sol advice score per advised action (Panel B of Figure 7); respondents at the median or without an advised action belong to neither group. Panel H: Wald p-values. Standard errors in parentheses. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table A15 Who Receives Which Spending Advice

	(1) Advice score	(2) Advice score	(3) Advice score	(4) Advice score
Female	-0.351***	-0.355***	-0.359***	-0.370***
	(0.092)	(0.092)	(0.093)	(0.091)
High household income	0.417***	0.432***	0.430***	0.461***
	(0.105)	(0.105)	(0.105)	(0.103)
Log prior non-durable spending	0.103**	0.110**	0.117**	0.116**
	(0.052)	(0.052)	(0.052)	(0.051)
Age 55+	0.600***	0.622***	0.620***	0.550***
	(0.120)	(0.120)	(0.120)	(0.118)
College degree	-0.032	-0.034	-0.033	0.014
	(0.098)	(0.099)	(0.099)	(0.097)
Employed	-0.097	-0.079	-0.067	-0.091
	(0.119)	(0.119)	(0.120)	(0.116)
Democrat		-0.121	-0.119	-0.091
		(0.109)	(0.109)	(0.106)
Republican		-0.374***	-0.359***	-0.353***
		(0.129)	(0.130)	(0.128)
Other party / prefer not to say		0.154	0.141	0.339
		(0.273)	(0.273)	(0.260)
Prior AI use: occasional			-0.057	-0.029
			(0.130)	(0.126)
Prior AI use: daily or more			-0.158	-0.120
			(0.162)	(0.156)
Number of turns				-0.242***
				(0.024)
Info + Chatbot arm	-0.052	-0.057	-0.058	-0.072
	(0.092)	(0.092)	(0.092)	(0.090)
Joint test: demographics (p)	0.000	0.000	0.000	0.000
Joint test: party (p)		0.021	0.033	0.011
Joint test: prior AI use (p)			0.595	0.691
Mean of dependent variable	-2.21	-2.21	-2.21	-2.21
Observations	2,283	2,283	2,283	2,283
R^2	0.025	0.030	0.030	0.075

Notes: OLS of the whole-conversation advice score; negative values mean net advice to spend less. All columns control for the Info + Chatbot arm. Robust standard errors in parentheses. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table A16 Requests and the Chatbot's First Reply

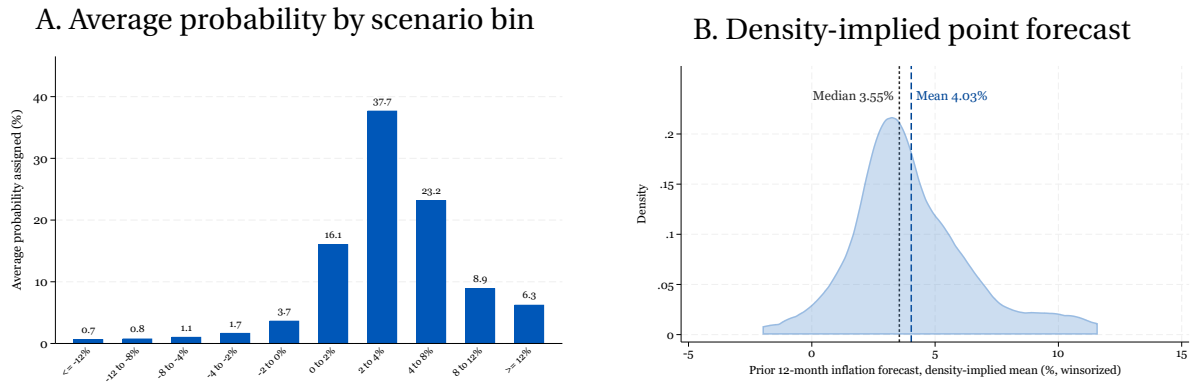
	(1)	(2)	(3)	(4)	(5)
Panel A. Advice score of the chatbot's first reply					
Female	-0.113**	-0.121***	-0.123***	-0.120***	-0.063
	(0.046)	(0.046)	(0.046)	(0.046)	(0.043)
High household income	0.199***	0.203***	0.202***	0.201***	0.166***
	(0.051)	(0.052)	(0.052)	(0.052)	(0.049)
Log prior non-durable spending	0.040	0.044	0.047*	0.047*	0.027
	(0.027)	(0.027)	(0.027)	(0.027)	(0.026)
Age 55+	0.142**	0.156**	0.155**	0.163**	0.068
	(0.063)	(0.063)	(0.063)	(0.064)	(0.060)
College degree	0.026	0.020	0.020	0.022	0.050
	(0.048)	(0.049)	(0.049)	(0.048)	(0.046)
Employed	-0.021	-0.011	-0.006	-0.004	-0.003
	(0.059)	(0.060)	(0.060)	(0.060)	(0.055)
Democrat		0.026	0.027	0.024	0.014
		(0.055)	(0.055)	(0.055)	(0.051)
Republican		-0.132**	-0.125*	-0.128*	-0.076
		(0.065)	(0.066)	(0.066)	(0.060)
Other party / prefer not to say		0.105	0.100	0.103	0.059
		(0.120)	(0.120)	(0.120)	(0.113)
Prior AI use: occasional			-0.012	-0.008	0.006
			(0.063)	(0.063)	(0.060)
Prior AI use: daily or more			-0.069	-0.069	-0.011
			(0.077)	(0.077)	(0.073)
First-message length (100 characters)				-0.064	0.068
				(0.040)	(0.057)
Info + Chatbot arm	0.040	0.039	0.039	0.036	0.054
	(0.046)	(0.046)	(0.046)	(0.046)	(0.043)
First-message codes	No	No	No	No	Yes
Joint test: demographics (p)	0.000	0.000	0.000	0.000	0.000
Joint test: party (p)		0.036	0.054	0.052	0.368
Joint test: first-message codes (p)					0.000
Mean of dependent variable	-0.74	-0.74	-0.74	-0.74	-0.74
Observations	2,282	2,282	2,282	2,282	2,282
R^2	0.015	0.019	0.019	0.021	0.196

	(1) Own question	(2) Budgeting / coping	(3) Plans to cut	(4) Discloses context	(5) Inflation itself
Panel B. Content of the respondent's first message (linear probability models)					
Female	0.009	0.028**	0.016**	0.026*	-0.033***
	(0.020)	(0.013)	(0.007)	(0.015)	(0.013)
High household income	-0.030	-0.034**	-0.000	0.007	-0.027**
	(0.023)	(0.015)	(0.009)	(0.017)	(0.014)
Log prior non-durable spending	0.015	0.009	-0.003	0.007	0.007
	(0.012)	(0.008)	(0.004)	(0.009)	(0.007)
Age 55+	0.148***	-0.019	0.011	0.121***	0.034*
	(0.030)	(0.018)	(0.012)	(0.025)	(0.019)
College degree	0.012	0.015	0.011	-0.002	0.007
	(0.022)	(0.014)	(0.008)	(0.016)	(0.013)
Employed	-0.041	0.008	-0.013	-0.019	0.006
	(0.027)	(0.017)	(0.011)	(0.020)	(0.016)
Democrat	-0.041*	-0.018	-0.018*	-0.014	-0.005
	(0.025)	(0.016)	(0.010)	(0.018)	(0.015)
Republican	-0.061**	0.005	-0.019*	-0.025	-0.029*
	(0.029)	(0.019)	(0.011)	(0.020)	(0.017)
Other party / prefer not to say	0.025	-0.026	-0.013	0.038	-0.032
	(0.060)	(0.036)	(0.020)	(0.046)	(0.034)
Prior AI use: occasional	0.012	0.010	0.031***	0.016	-0.017
	(0.030)	(0.018)	(0.008)	(0.021)	(0.019)
Prior AI use: daily or more	-0.109***	-0.004	0.009	-0.019	-0.052**
	(0.035)	(0.022)	(0.009)	(0.025)	(0.021)
Info + Chatbot arm	-0.040**	-0.010	-0.001	-0.036**	-0.010
	(0.020)	(0.013)	(0.007)	(0.015)	(0.012)
Joint test: demographics (p)	0.000	0.081	0.183	0.000	0.025
Joint test: party (p)	0.116	0.436	0.261	0.435	0.257
Joint test: prior AI use (p)	0.000	0.668	0.000	0.130	0.020
Mean of dependent variable	0.39	0.11	0.03	0.15	0.10
Observations	2,282	2,282	2,282	2,282	2,282
R^2	0.030	0.006	0.012	0.023	0.011

Notes: Panel A: OLS of the first reply's advice score; column 5 adds dummies for the coded content of the first message. Panel B: linear probability models of first-message content. All columns control for the Info + Chatbot arm. Robust standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

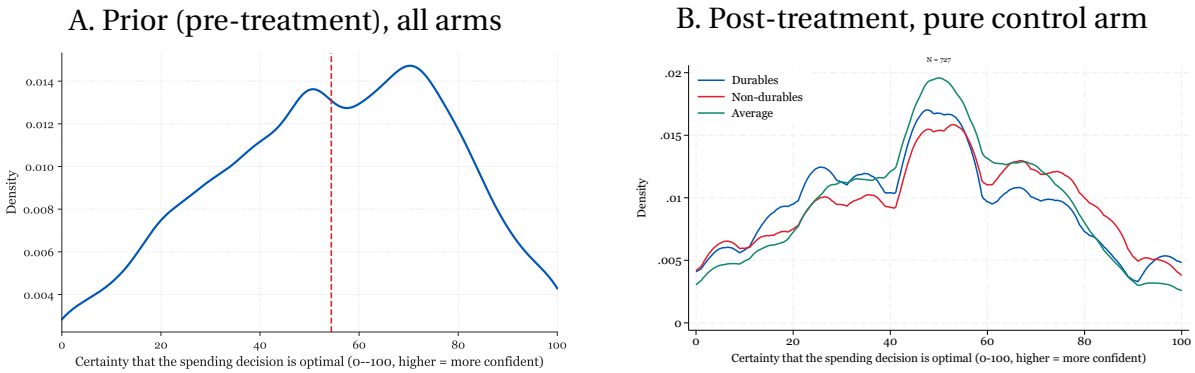
A.2 Additional Figures

Figure A1 Prior Inflation Beliefs



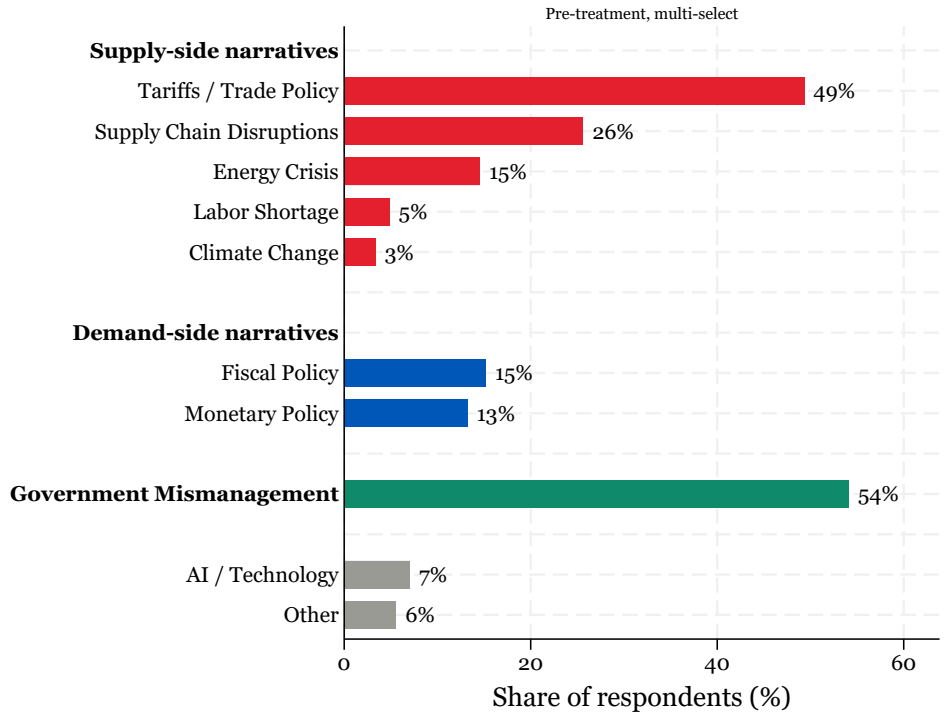
Notes: Analysis sample (N = 3,870). Panel A: average probability assigned to each scenario bin. Panel B: kernel density of each respondent's density-implied prior 12-month inflation forecast (winsorized 1/99); vertical lines mark the median (dotted) and mean (dashed).

Figure A2 Decision Confidence: Prior and Pure Control Arm



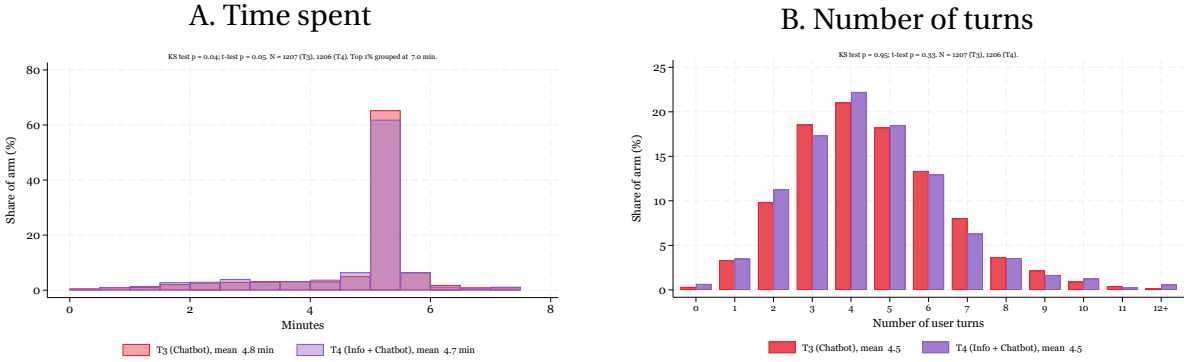
Notes: Kernel densities of the decision-confidence slider (0–100). Panel A: pre-treatment slider, all arms (N = 3,870). Panel B: post-treatment sliders for durable and non-durable spending, T1 only (N = 727). Dashed vertical lines mark the means.

Figure A3 Prior Driver of Inflation Forecast



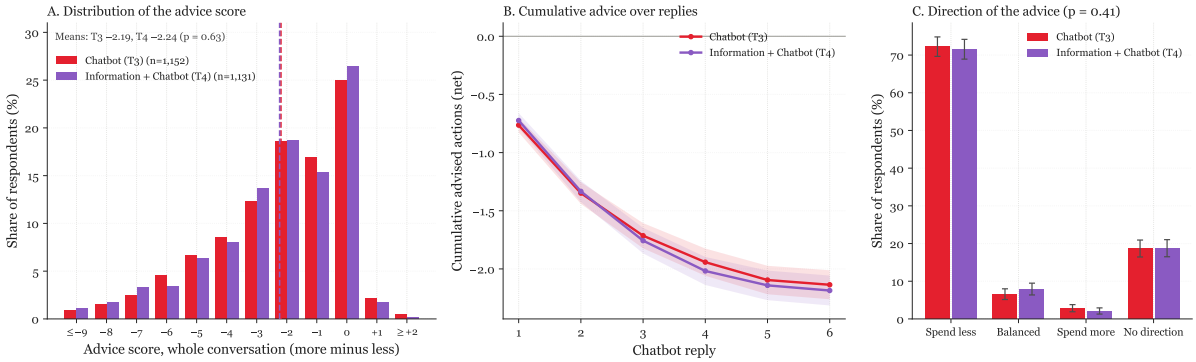
Notes: Share of respondents selecting each factor in the pre-treatment multi-select inflation-driver question (N = 3,870). Respondents may select more than one factor.

Figure A4 Chatbot Engagement, T3 (Chatbot) vs. T4 (Information + Chatbot)



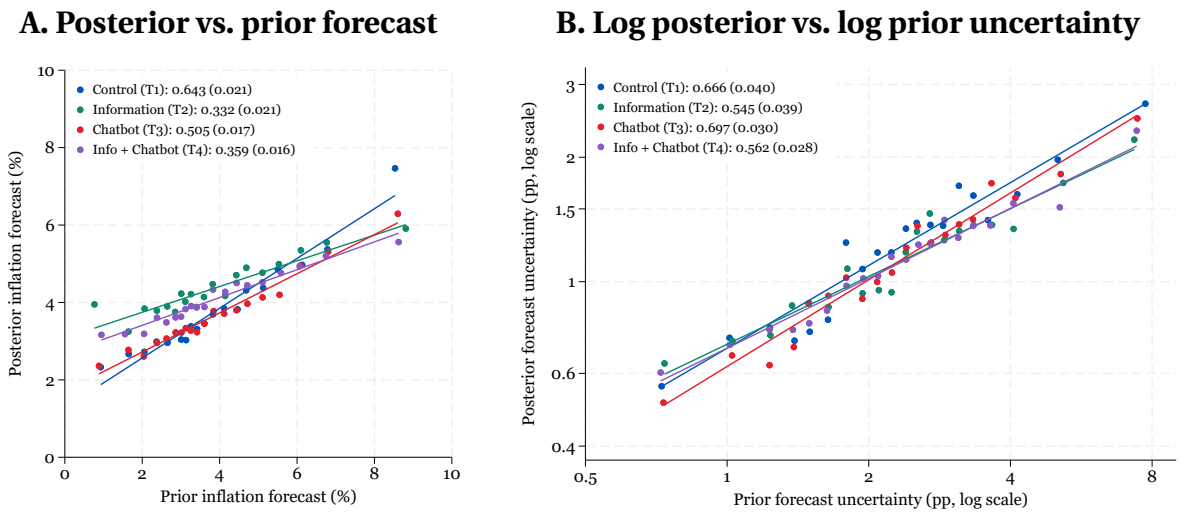
Notes: Chatbot-arm respondents by arm. Panel A: minutes on the chat page, top 1% grouped. Panel B: user messages sent, excluding the scripted opening message, top-coded at 12. Subtitles report two-sample Kolmogorov-Smirnov and t-test p-values. N = 1,207 (T3), 1,206 (T4).

Figure A5 Spending Advice by Treatment Arm



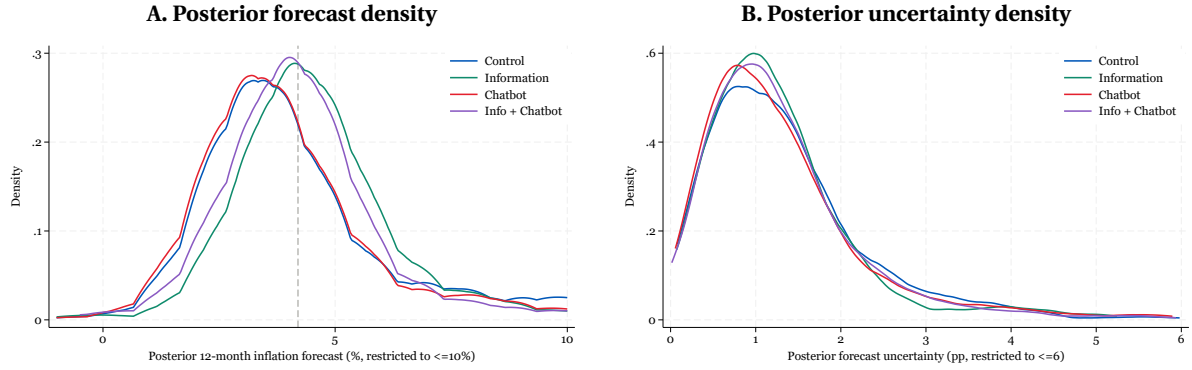
Notes: GPT-6.1 Sol coding. Panel A: distribution of the whole-conversation advice score. Panel B: mean cumulative advice score after each of the first six replies; values are carried forward after a conversation ends. Panel C: direction of the advice. $N = 1{,}152$ (T3), 1,131 (T4).

Figure A6 Updating of the Inflation Forecast and Uncertainty, by Arm



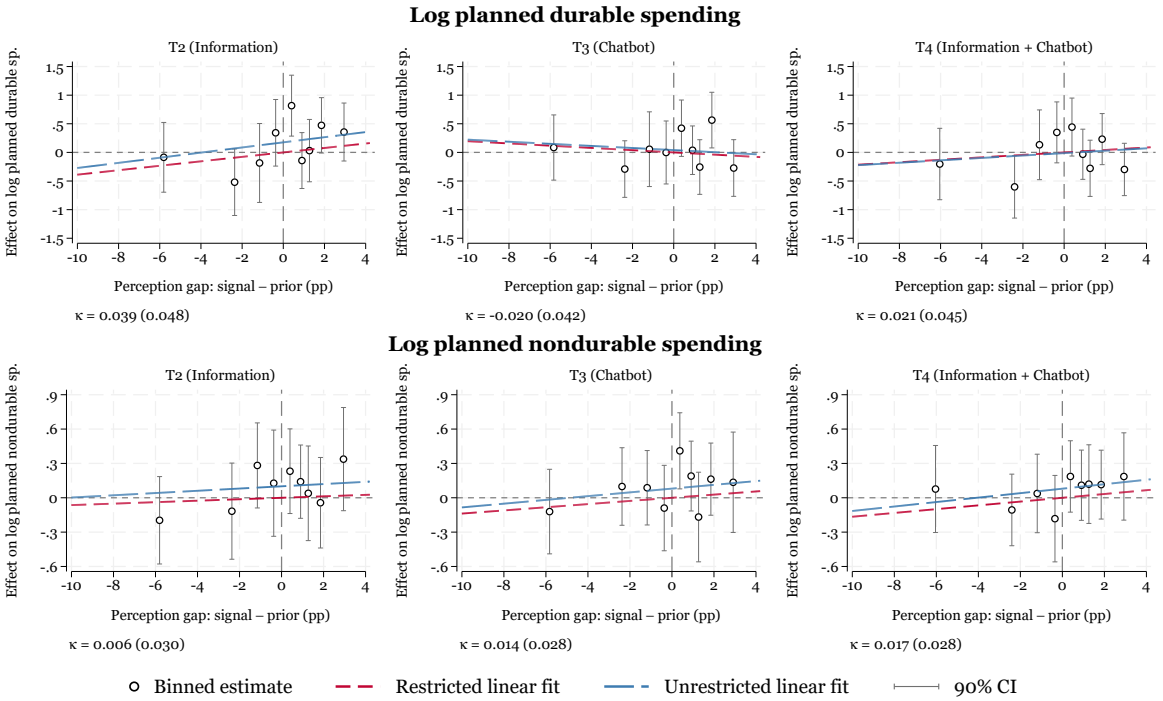
Notes: Panel A: binned scatter of the posterior on the prior 12-month inflation forecast, with a separate OLS fit by arm (slope, robust SE in parentheses); $N = 3{,}870$. Panel B: binned scatter of log posterior on log prior forecast uncertainty, with a separate weighted OLS fit by arm, excluding the 170 respondents with zero prior uncertainty. Axis labels give uncertainty in percentage points on a log scale.

Figure A7 Posterior Inflation Forecast and Forecast Uncertainty: Distributions by Arm



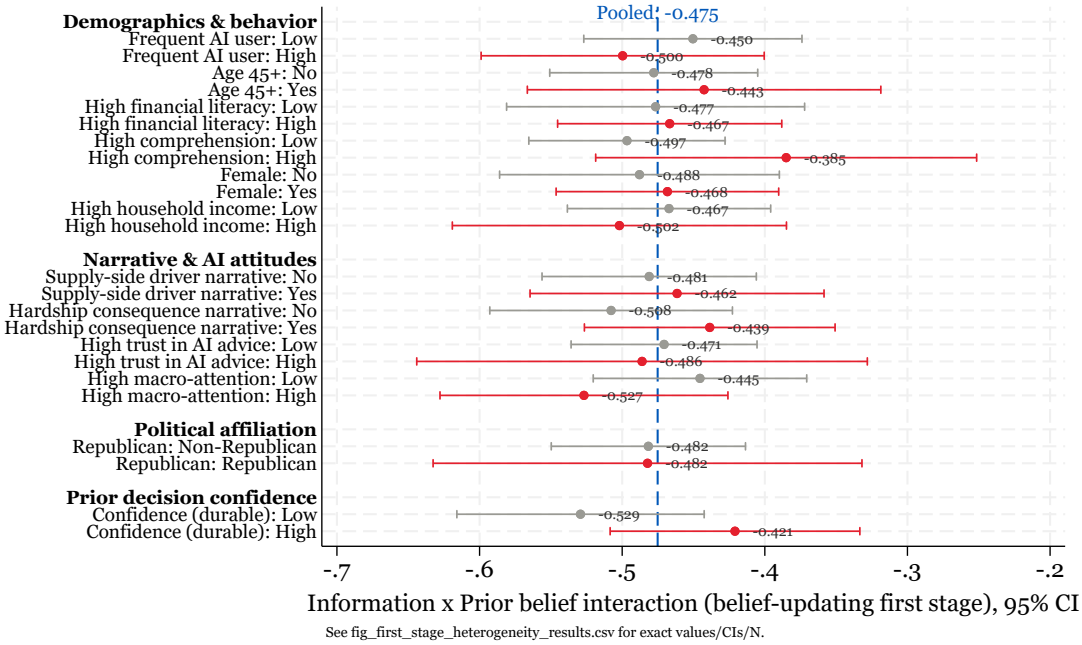
Notes: Kernel densities by arm. Panel A: posterior 12-month inflation forecast, restricted to ≤ 10%; the dashed line marks the 4.2% CPI figure. Panel B: posterior forecast uncertainty, restricted to ≤ 6 percentage points.

Figure A8 Effect of Each Treatment on Planned Spending, by Perception Gap



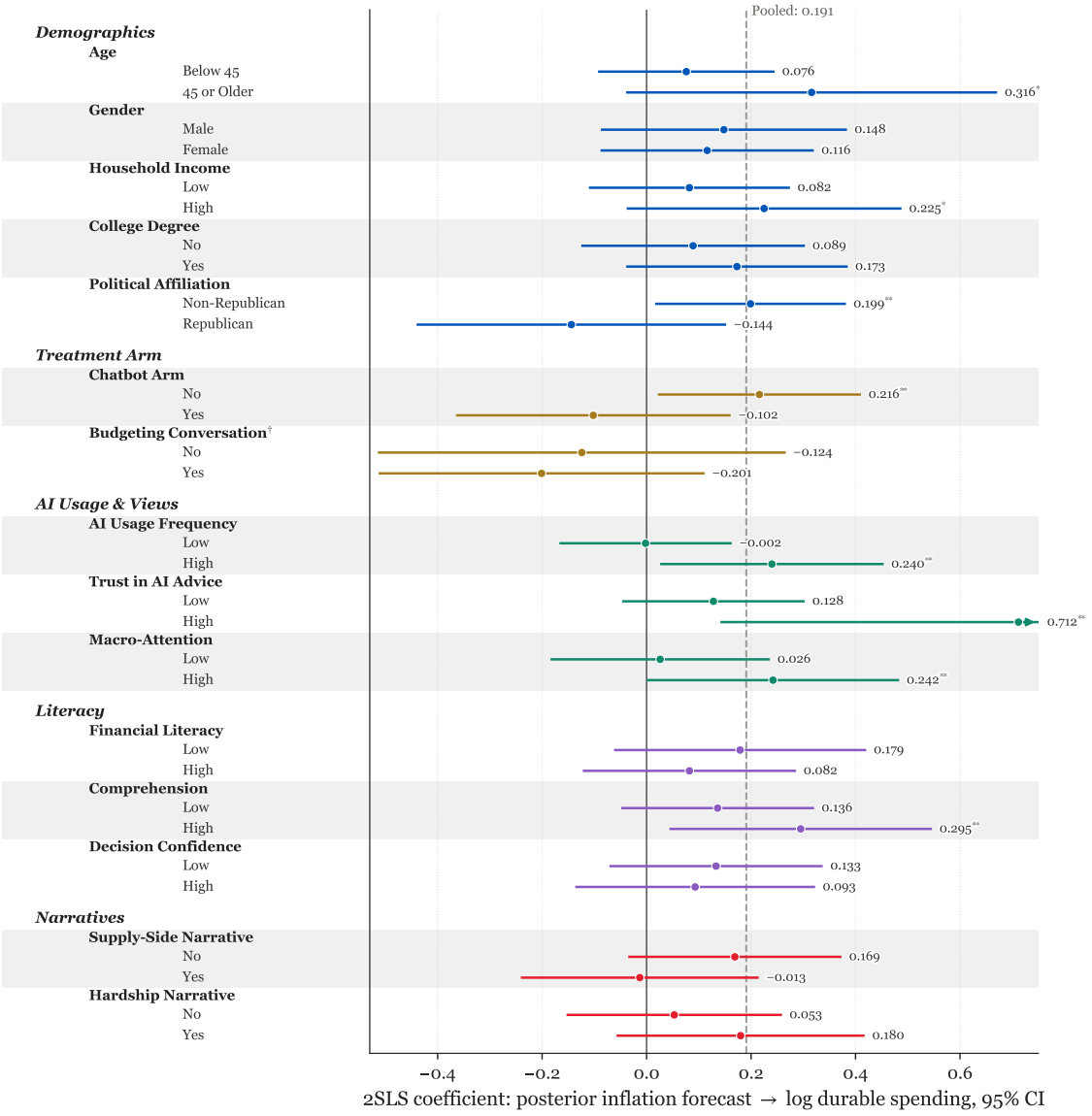
Notes: Each column compares one arm with the control group, among respondents with a positive density-implied prior: T2 (N = 1,394), T3 (N = 1,852), T4 (N = 1,840). The perception gap is the 4.2% signal minus the prior mean. Circles: coefficient on the arm dummy from OLS within each gap bin, with 90% confidence intervals (HC1). Dashed red line: restricted linear fit $\tau(g) = \kappa g$. Long-dashed blue line: unrestricted fit $\tau(g) = d + eg$. All regressions control for female, college degree, high household income, employed, age-band dummies, and the outcome's pre-treatment counterpart.

Figure A9 Heterogeneity in the Belief-Updating First Stage: the Information × Prior Interaction



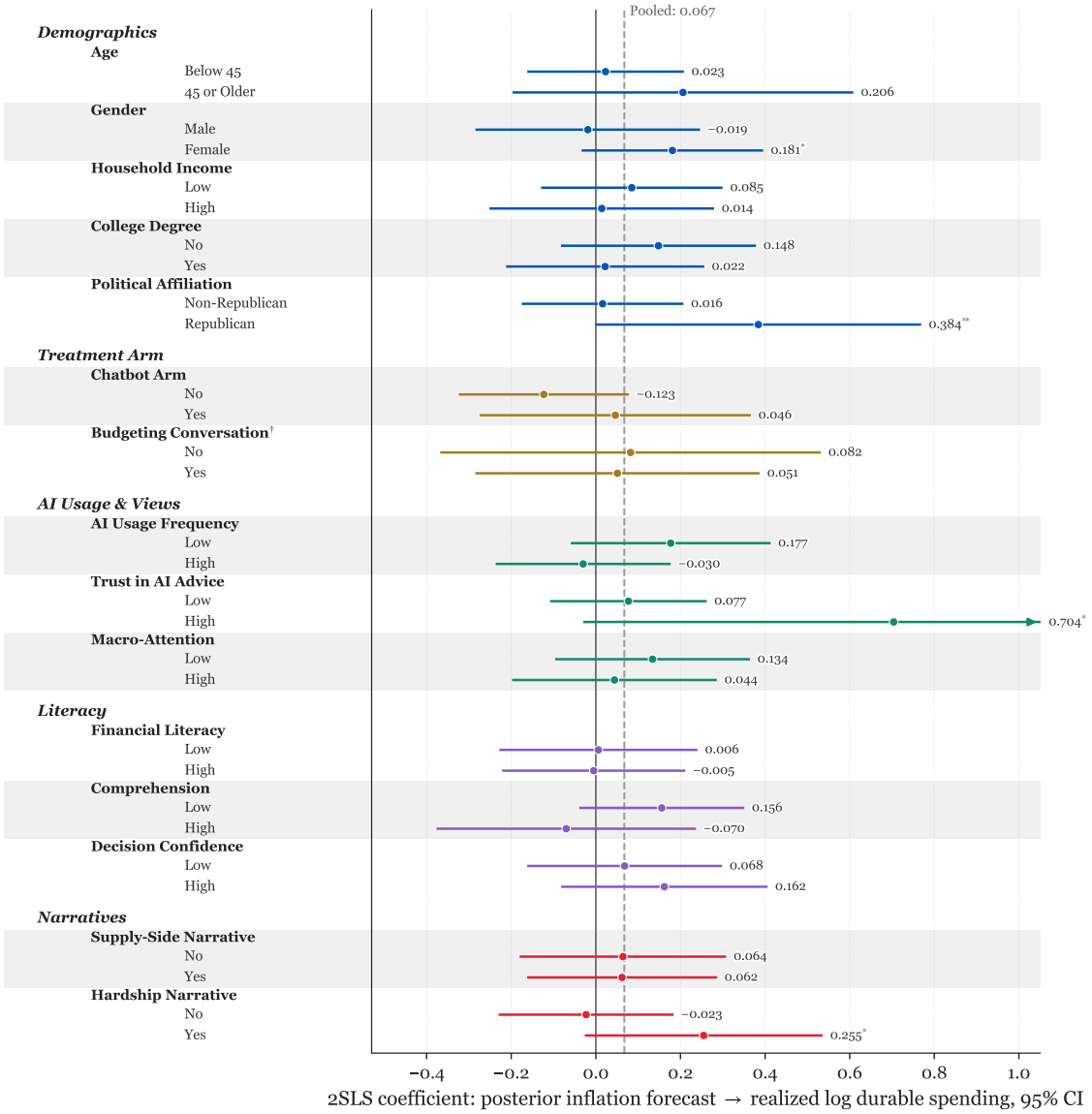
Notes: First stage of column (1) of Table 6, re-estimated on one half of the sample at a time, split by the named moderator. Each row plots the T2 × Prior interaction, with 95% confidence intervals. The dashed line marks the pooled estimate.

Figure A10 Heterogeneity in the Durable-Spending Pass-Through, Jackknife-Trimmed



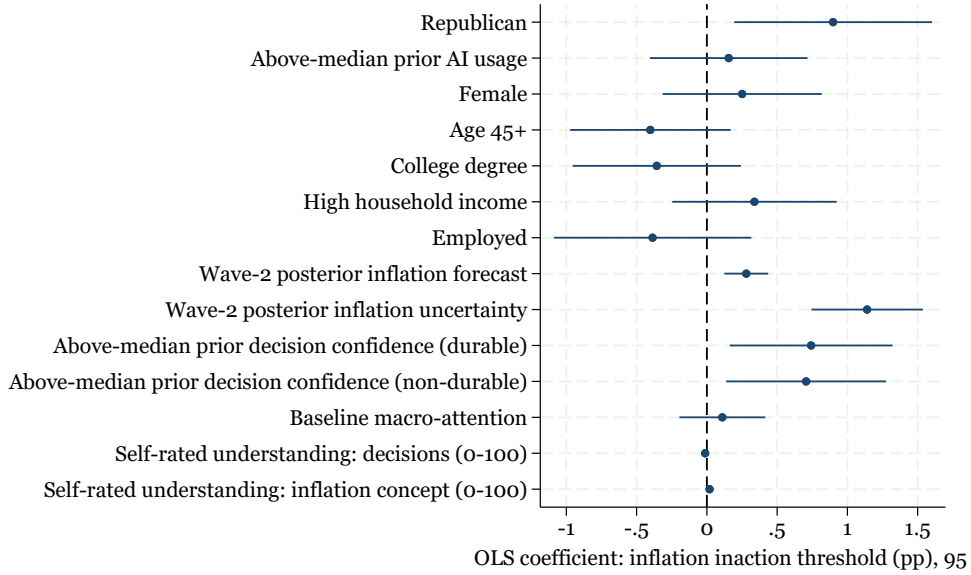
Notes: Column (1) of Table 6, re-estimated on one half of the sample at a time after a jackknife trim within that subsample. 95% confidence intervals. The dashed line marks the pooled estimate; an arrowhead marks an interval cut off at the axis. †Defined only within T3 and T4. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure A11 Heterogeneity in the Realized Durable-Spending Pass-Through, Jackknife-Trimmed



Notes: Realized-spending analogue of Figure A10: the outcome is realized log durable spending in the follow-up, follow-up respondents only, with days between the two surveys as an additional control. 95% confidence intervals. The dashed line marks the pooled estimate; an arrowhead marks an interval cut off at the axis. †Defined only within T3 and T4. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Figure A12 Determinants of Inflation Inaction Band



Notes: OLS, robust (HC1) standard errors. The outcome is the inflation inaction threshold (percentage points, winsorized 1/99). Regressors are shown in the figure; decision confidence enters as above-median dummies, and self-rated understanding as a percentile (0–100). Follow-up respondents, $N = 3{,}074$. Whiskers: 95% confidence intervals.

B Survey Transcripts

B.1 Main Survey

We provide the main survey instructions below. Text shown in *italics* is not shown to respondents.

Introduction

Information sheet and informed consent form (IRB-approved). Respondents who do not consent are routed to early termination and are not paid.

I consent to the form above and wish to participate in the study. [Yes; No]

Demographics

In the following, we would like to ask you a few questions about your background.

What is your **age**? [18-24; 25-34; 35-44; 45-54; 55-64; 65 or above]

[Page break]

What is your **gender**? [Male; Female; Other; Prefer to self-describe:]

[Page break]

What is the **highest level of education** you have completed? [Less than high school; High school diploma or equivalent; Some college, no degree; Associate's degree; Bachelor's degree; Master's degree; Doctoral degree (PhD, MD, JD, etc.); Prefer not to answer]

[Page break]

What is your approximate **annual household income** before taxes? [Less than $25,000; $25,000 – $49,999; $50,000 – $74,999; $75,000 – $99,999; $100,000 – $149,999; $150,000 – $199,999; $200,000 or more; Prefer not to answer]

[Page break]

What is your current **employment status**? [Employed full-time; Employed part-time; Self-employed; Unemployed and looking for work; Not in the labor force (student, retired, homemaker, etc.); Prefer not to answer]

[Page break]

What is your **political affiliation**? [Democrat; Republican; Independent; Other; Prefer not to answer]

[Page break]

What was your **household's average monthly spending on nondurable goods and services** over the last three months? *Nondurable goods and services include, for instance, food, tobacco, alcohol, gasoline,*

clothing, haircuts, transportation, and other small services and nondurable goods that do not last over time. Please provide an answer in dollars. If unsure, please estimate. dollars per month: _____

AI Usage

In this section we would like to ask you some questions about your **use of AI assistants (e.g., ChatGPT, Claude, Gemini) to inform yourself about economic topics.**

How often, if ever, have you used an **AI assistant (e.g., ChatGPT, Claude, Gemini) specifically to get information or form a view about economic topics** — such as prices and the cost of living, the job market, interest rates, or your personal financial situation? [Several times a day; About once a day; A few times a week; A few times a month; Rarely (less than once a month); Never]

[Page break]

[Shown only to respondents who did not select "Never" in the first question of this section.]
For **which of the following economic topics**, if any, **have you used an AI assistant in the past 3 months** to inform yourself? Please select all that apply. *[Select all that apply]*
[Future prices / cost of living (inflation); Job market conditions (availability of jobs, wages); Interest rates (mortgages, savings rates, credit); Stock market or investment returns; Your own household income or earnings outlook; Government policy (taxes, benefits, regulations); General economic growth or recession risk; None of these]

[Page break]

[Shown only to respondents who did not select "Never" in the first question of this section.]
Still thinking about when you have used an **AI assistant for economic topics**: which of the following best describes how you **typically used it**? Please select all that apply. *[Select all that apply]*
[To understand a concept or term (e.g., "What does the inflation rate measure?"); To get an actual figure or forecast (e.g., "What is the expected inflation rate this year?"); To get an explanation of why something is happening in the economy; To understand how economic conditions might affect my personal situation; To check or challenge something I read or heard elsewhere; To help make a specific financial decision (e.g., whether to fix a mortgage rate or make major purchases); Other (please specify):]

[Page break]

Compared to each of the following, how accurate do you think **information from an AI assistant** is when it comes to **economic topics** like prices, jobs, or interest rates?

- News media
- Financial professional
- Government / central bank
- Family and friends

[Response options: Much less accurate, Somewhat less accurate, Similarly accurate, Somewhat more accurate, Much more accurate, Don't know]

[Page break]

For each of the following situations, please indicate how much you would **trust information from an AI assistant** to help you form a view.

- Understanding a general concept (e.g., how inflation is measured)
- Getting a current factual figure (e.g., last month's inflation rate)
- Getting a forecast for the next 12 months (e.g., expected inflation)
- Assessing how a policy change will affect the economy
- Deciding whether to fix or float a mortgage rate
- Deciding whether to change my investments or savings

[Response options: Not at all, Slightly, Moderately, Quite a lot, Completely]

[Page break]

[Shown only to respondents who did not select "Never" in the first question of this section.]
Think of the most recent time you used an **AI assistant to get information about an economic topic**. Did the information or explanation you received **change your view or expectation** about that topic? Please answer on a scale from 0 (your view did not change) to 10 (your view completely changed).
[Slider (How much did the information from the AI assistant change your view?), 0 to 10]

Macroeconomic Information Acquisition

We would like to know how you typically stay informed about the macroeconomy.

How often do you typically **acquire information about the macroeconomy (e.g., inflation, interest rates, economic growth, unemployment)** from the **following sources**?

- TV, radio, newspapers
- Online news websites (e.g., NYT, FT, Bloomberg, WSJ)
- Social media (e.g., Twitter/X, Instagram, Facebook, Reddit)
- Podcasts or YouTube channels on economics/finance
- AI assistants such as ChatGPT, Claude
- Official government or statistical agency websites (e.g., BLS, Fed)
- Friends, family, or colleagues

- My own professional experience
- Financial advisors or economic experts

[Response options: Daily, Several times a week, A few times a month, Once a month, Once in a few months, Once a year, Never]

Inflation Expectations and Narratives Prior Elicitation

In the following, we would like to ask you about your **expectations for inflation** and your views on its causes.

Inflation Some of the following questions are about your views on and beliefs about the inflation rate in the United States. Here you are provided with a **definition of the term inflation**. Please read it carefully. You can proceed after a certain minimum time. Inflation measures how much prices in the economy rise from year to year. It is defined as the **yearly increase in the general price level of goods and services** (Consumer Price Index). If the inflation rate is **negative**, this is called **deflation**, meaning that prices decrease from one year to the next.

Hypothetically, if the **annual inflation rate in the United States over the past 12 months was 2.7%,** a typical bundle of goods and services that did cost $1,000 a year ago, would cost $1,027 now. *[An image is displayed next to the example.]*

[Page break]

In this question, we present you with possible scenarios for the **inflation rate in the United States over the next 12 months**. Please let us know how likely you think it is that each scenario will occur. Type a number to indicate the probability, in percent, that you attach to each scenario. The probabilities must sum to 100 percent.

Inflation will be...

- 12% or higher: _____
- between 8% and 12%: _____
- between 4% and 8%: _____
- between 2% and 4%: _____
- between 0% and 2%: _____
- between -2% and 0% (deflation of 0% to 2%): _____
- between -4% and -2% (deflation of 2% to 4%): _____
- between -8% and -4% (deflation of 4% to 8%): _____
- between -12% and -8% (deflation of 8% to 12%): _____

- -12% or lower (deflation of 12% or more): _____

[Answers must sum to 100]

[Page break]

In your opinion, what has been the **most important driver of current inflation**? Please select up to two drivers. *[Select up to two]*
[Fiscal Policy; Monetary Policy; Tariffs; Energy Crisis; Labor Shortage; Climate Change; AI; Supply Chain Disruptions; Government Mismanagement; Other:]

[Page break]

Do you plan to **buy any of the following goods over the next month**? Please select all that apply. *[Select all that apply]*
[Car, truck, or other vehicle; House; Major durable good (washing machine, TV, etc.); Boat; Luxury items; None of the above]

[Page break]

How **confident** would you be to make **optimal spending decisions** based on your **inflation expectations**? *[Slider (Confidence in %), 0 to 100]*

WTP for Different Explanations

Advice on Inflation
Suppose you receive **news about inflation**. You may get **advice from various sources on what it means for your spending** — for example, whether it is a good time to buy durable goods, delay purchases, or adjust your household budget. On the next screen, we will ask you to **rank the following sources** from most to least preferred for this type of advice:

- **AI Assistant**: an AI chatbot trained on a frontier large language model.
- **Economic Experts**: economists with expertise on inflation and household decision-making.
- **U.S. Federal Reserve**: information from the Federal Reserve, the central bank of the United States.
- **Media**: information from a newspaper article or a TV program.
- **Social Circle**: friends, family, colleagues

Please rank the source first you would prefer the **most**.

[Page break]

From **which source** would you prefer to receive **information about the implications of inflation for spending**? Drag to arrange the sources from highest preference (top of the list) to the lowest preference (bottom of the list). *[Drag-and-drop ranking, most to least preferred:]*

- AI Assistant
- Economic Experts
- U.S. Federal Reserve
- Media
- Social Circle

[Page break]

Given current conditions, what is the **maximum amount you would be willing to pay** to receive advice from the **following sources**? Please enter a value in dollars.

- AI Assistant: _____
- Economic Experts: _____
- U.S. Federal Reserve: _____

Attention Check

The next question is about the following problem. In questionnaires like ours, sometimes there are participants who do not carefully read the questions and just quickly click through the survey. This compromises the results of research studies. To show that you are reading the survey carefully, please select "Orange" as your answer to the below question.

Given the above, what is the color of the sky? [Red; Blue; Green; Orange; Black]

[Respondents who do not select "Orange" are routed to early termination.]

Inflation Treatment

[Shown only to respondents randomly assigned to the information-treatment arms (T2 and T4; Section 2.2).]

Information on Current Inflation
On the next screen, you will see information on the **current inflation** in the United States. **Please read that information carefully** before you proceed.

[Page break]

Inflation in the United States
We would like you to know that, according to the U.S. Bureau of Labor Statistics, the **Consumer Price Index in the United States increased by 4.2% over the last 12 months**.

Chatbot Interaction

[Shown only to respondents randomly assigned to the chatbot-treatment arms (T3 and T4; Section 2.2).]

Use an AI Assistant for Inflation & Spending Questions
On the next screen, you will be able to ask an **AI assistant** questions about **inflation in general** and **its implications for your spending and saving decisions**. This assistant is based on a frontier LLM. Your questions will be coded after the survey and the **20% of respondents with the highest score** for **engagement and fit with the topic** will receive an additional **$1**. You can use the assistant **for up to 5 minutes**.

[Page break]

Embedded chat interface with the AI assistant. Respondents may type their messages or use a voice-recording feature (buttons "Record response" and "Submit response"); the note "Please write full sentences." and a five-minute countdown timer ("Time remaining") are displayed. The assistant is built on a frontier large language model (Section 2.3).

[Page break]

How **helpful** was the **interaction with the AI assistant**? *[Slider (Please answer on a scale from 1 to 10), 1 to 10]*

[Page break]

Do you feel like you **learned something** from the interaction? [Yes; No]

[Page break]

[Shown only if the respondent answered "Yes" to the question above.]
What did you learn from the interaction? State the most important learnings.

_____

Posteriors 1 - Expectations and Narratives

Inflation Expectations and Narratives
We would now like to ask you again about your **inflation expectations** and your views on its **consequences for households**.

What do you think the **inflation rate** (as measured by the Consumer Price Index) **is going to be over the next 12 months**? Please provide your answer as a percentage change from current prices. If you expect inflation, enter a positive number. If you expect deflation, enter a negative number. If you expect neither, enter zero. Inflation over the next 12 months (in %): _____

[Page break]

Think about three **plausible scenarios for inflation**. Start with the LOWEST inflation in general and end with the HIGHEST inflation over the next 12 months. What do you think will be the approximate

inflation in general for each of the scenarios? Lowest inflation (in %): _____ Medium inflation (in %): _____ Highest inflation (in %): _____

[Page break]

Now we ask you to think about the **chance of the inflation scenarios you entered in the previous screen actually happening over the next 12 months**. Please assign a percentage chance to each of the inflation scenarios you entered to indicate how likely you think it is that this inflation scenario will actually happen over the next 12 months. Your answers can range from 1 to 100, where 1 means there is almost no chance that this price change will happen, and 100 means that it is absolutely certain that the price change will happen. The sum of the points you allocate should total to 100.

- Low inflation *respondent's own low-inflation scenario value, entered above*%: _____
- Medium inflation *respondent's own medium-inflation scenario value, entered above*%: _____
- High inflation *respondent's own high-inflation scenario value, entered above*%: _____

[Answers must sum to 100]

[Page break]

In your opinion, what is the **most important driver of current inflation**?

_____

[Page break]

What do you expect will happen to each of the **following indicators** in the United States over the next 12 months?

- Interest Rates
- Unemployment
- Inflation

[Response options: Decrease a lot, Decrease a little, Stay about the same, Increase a little, Increase a lot]

[Page break]

In your opinion, how should the **government respond to inflation**?

_____

[Page break]

How do you expect your **household's financial situation** to change over the next 12 months? It will... [get much worse; get somewhat worse; stay about the same; get somewhat better; get much better]

[Page break]

In your opinion, what are the **main economic consequences of inflation for your household**? Please try to answer in 2-3 sentences. _____

Posteriors 2 - Actual Spending

Spending Decisions
Please now think about how you expect your **spending** to **change** over the coming weeks and months.

Compared to *June*, how do you expect your **consumption of the following categories** to change in *July*?

- Durable goods
- Food at home
- Restaurants
- Non-durable goods
- Leisure

[Response options: Consume much less, Consume less, Consume about the same, Consume more, Consume much more]

[Page break]

How much do you **plan to spend on durable goods over the next 12 months?** Please enter an amount in dollars. _____

[Page break]

How **certain** are you that your **planned spending on durable goods is optimal given your expectations about future inflation?** *[Slider (Certainty (in %)), 0 to 100]*

[Page break]

How much do you **plan to spend on non-durable goods per month over the next 12 months?** Please enter an amount in dollars. _____

[Page break]

How **certain** are you that your **planned spending on non-durable goods is optimal given your expectations about future inflation?** *[Slider (Certainty (in %)), 0 to 100]*

[Page break]

Suppose you received an **unexpected payment of $10,000 today**. Please indicate how much you would **save or invest in each of the following asset categories**. Your answers should add up to $10,000.

- Savings or current accounts: _____
- Stocks and shares: _____
- Mutual funds and collective investments: _____
- Retirement or pension products: _____
- Bonds: _____
- Gold or other precious metals: _____
- Bitcoin and/or other crypto assets: _____

[Answers must sum to 10000]

Posteriors 3 - Hypothetical Spending

[The direction of the hypothetical shock (higher vs. lower) is randomized at the respondent level via a separate randomizer, independent of the information/chatbot treatment assignment.]

Hypothetical Inflation Shock
In the following we want you to think about a **purely hypothetical shock to inflation**, and elicit your potential **reactions** to it.

Hypothetically suppose that inflation over the next 12 months will be **2 percentage points *higher/lower, randomly determined*** than **your expectation** of ***respondent's own 12-month inflation point forecast, entered above*%.** In that case, how would you **adjust** the **quantity** of goods and services that you purchase relative to your current plan? [Increase; Unchanged; Decrease; Don't know]

[Page break]

[Shown only if the respondent answered "Increase" or "Decrease" to the question above.]
By **how much** would you ***increase/decrease, matching the respondent's answer above*** **the quantity of goods and services you purchase**? Please answer in %: _____

[Page break]

Again assume that **inflation over the next 12 months will be 2 percentage points *higher/lower, randomly determined* than your expectation of *respondent's own 12-month inflation point forecast, entered above*%.** In that case, how would you **adjust your behavior** in response, relative to your current plan? If inflation is **2 percentage points *higher/lower, randomly determined*** than I expected, I will ... *[Five of the thirteen statements below are shown, selected at random.]*

- Bring forward major purchases of durable goods
- Reduce usual spending and put aside more money
- Shop around more actively to find the best price for the same exact product or service

- Substitute goods and services with cheaper alternatives
- Switch stores
- Shop more online
- Save less than usual or liquidate (some or all) savings to finance spending
- Use more credit than usual to finance spending (e.g. overdrawing a checking account, taking on more credit card debt, or other consumer loans)
- Buy gold, real estate and other inflation-protected assets
- Hold more savings in cash
- Adopt a less risky portfolio strategy
- Ask for a pay raise or look for a higher paying job
- Look to increase my income in other ways (e.g. take on a second job, work more hours)

[Response options: Much less likely, Less likely, Similar, More likely, Much more likely]

Comprehension Checks

Concepts in Economics
We would like to ask you a few questions about **economic concepts**.

Suppose you expect **prices to rise significantly** over the next year. How does this affect the **real return on saving money**? [Higher expected inflation raises the real return on savings, making saving more attractive; Higher expected inflation lowers the real return on savings, making saving less attractive; Higher expected inflation has no effect on the real return on savings; I don't know]

[Page break]

If you expect **prices to be 5% higher a year from now**, what does this imply for the **best time to make a large purchase**, such as a new appliance or car? [It is better to wait and buy next year, because prices will be higher and the purchase is more important then; It does not matter when you buy — expected inflation does not affect the timing of purchases; It is better to buy sooner rather than later, because the same item will cost more in the future; I don't know]

[Page break]

Inflation is running at 5%. The **central bank raises its policy rate from 2% to 6% to fight inflation.** What is the approximate **real interest rate** now, and what does this imply for the **incentive to consume** today rather than save? [Real rate $\approx +11\%$; saving is very attractive, so consuming today is costly; Real rate $\approx +1\%$; there is now a small positive return to saving, slightly favoring deferring consumption; Real rate $\approx -3\%$; saving still has a negative real return, so consuming today remains preferable; I don't know]

Financial Literacy

Before you finish the survey, we will ask you a few additional questions about basic economic concepts.

Suppose you had $100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow? [More than $102; Exactly $102; Less than $102; I do not know; Prefer not to answer]

[Page break]

Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, would you be able to buy more than, exactly the same as, or less than today with the money in this account? [More than today; Exactly the same; Less than today; I do not know; Prefer not to answer]

[Page break]

Do you think the following statement is true or false? "Buying a single company's stock usually provides a safer return than a stock mutual fund." [True; False; I do not know; Prefer not to answer]

Comments

Thank you very much for participating in this survey! If you have any **comments or suggestions**, please write them here. _____

B.2 Follow-Up Survey

We provide the follow-up survey instructions below. Text shown in *italics* is not shown to respondents. *The survey was sent to a subset of main-survey respondents about one month after the main survey (Section 2.4).*

Introduction

Information sheet and informed consent form (IRB-approved), stating a survey length of about 7 minutes and a payment of $4. Respondents who do not consent are routed to early termination.

Inflation Expectations and Narratives

In the following, we would like to ask you about your expectations for inflation and your views on its causes.

How much do you **trust information from each of the following sources** about **economic topics** such as inflation, interest rates, and the job market?

- Federal Reserve / Central Bank
- Academic economists and financial experts
- News media (e.g., USA Today, New York Times Bloomberg, Wall Street Journal)
- Government statistical agencies (e.g., Bureau of Labor Statistics, Census Bureau)
- Financial professionals / advisors
- AI tools (e.g., ChatGPT, Claude, Gemini)

[Response options: Completely untrustworthy, Untrustworthy, Neither trustworthy nor untrustworthy, Trustworthy, Completely trustworthy; rows displayed in random order]

[Page break]

What do you think the **inflation rate** (as measured by the Consumer Price Index) **is going to be over the next 12 months**? Please provide your answer as a percentage change from current prices. If you expect inflation, enter a positive number. If you expect deflation, enter a negative number. If you expect neither, enter zero. Inflation over the next 12 months (in %): _____

[Page break]

Think about three **plausible scenarios for inflation**. Start with the LOWEST inflation in general and end with the HIGHEST inflation over the next 12 months. What do you think will be the approximate inflation in general for each of the scenarios? Lowest inflation %: _____ Medium inflation %: _____ Highest inflation %: _____

[Page break]

Now we ask you to think about the **chance of the inflation scenarios you entered in the previous screen actually happening over the next 12 months**. Please assign a percentage chance to each of the inflation scenarios you entered to indicate how likely you think it is that this inflation scenario will actually happen over the next 12 months. Your answers can range from 1 to 100, where 1 means there is absolutely no chance that this price change will happen, and 100 means that it is absolutely certain that the price change will happen. The sum of the points you allocate should total to 100.

- Low inflation *respondent's own low-inflation scenario value, entered above*%: _____
- Medium inflation *respondent's own medium-inflation scenario value, entered above*%: _____
- High inflation *respondent's own high-inflation scenario value, entered above*%: _____

[Answers must sum to 100]

[Page break]

In your opinion, what is the **most important driver of current inflation**?

_____

[Page break]

How **confident** would you be to make **optimal spending decisions** based on your **inflation expectations**? *[Slider (Confidence in %), 0 to 100]*

[Page break]

In your opinion, what are the **main economic consequences of inflation for your household**? Please try to answer in 2-3 sentences. _____

Realized Spending Behavior

Please now think about your recent spending behavior.

How much did your **consumption change** in **August relative to July** in each of the following categories?

- Food at home
- Food at restaurants
- Durable goods (e.g., appliances, electronics, furniture)
- Non-durable goods (e.g., clothing, personal care)
- Leisure (e.g., entertainment, travel, sports)

[Response options: Much less, Somewhat less, About the same, Somewhat more, Much more; rows displayed in random order]

[Page break]

How much did you actually **spend on durable goods** in **July**? Please enter an amount in dollars.

_____

[Page break]

How much did you actually **spend on non-durable goods** in **July**? Please enter an amount in dollars.

_____

[Page break]

In **July**, which of the **following did you do**? Please select all that apply. *[Select all that apply; options displayed in random order, except "None of the above", which is shown last]*
[Brought forward a major purchase to buy sooner than planned; Reduced usual spending and saved more; Shopped around more or switched stores to find better prices; Substituted goods or services with cheaper alternatives; Shopped more online; Saved less than usual or used existing savings to finance spending; Used more credit than usual (credit card debt, overdraft, etc.); Postponed or cancelled a planned large purchase; None of the above]

[Page break]

If you compare yourself to the rest of population, how well do you think you **understand the implications of inflation for households decisions**, like spending and saving decisions. I understand it better than.... *[Slider (Please answer in %), 0 to 100]*

Hypothetical Inflation Shock

[The direction of the hypothetical shock (higher vs. lower) is randomized at the respondent level via a separate randomizer with equal probabilities. The direction is drawn anew in the follow-up, independently of the main-survey draw.]

In the following we want you to think about a **purely hypothetical shock to inflation**, and elicit your potential **reactions** to it.

Hypothetically suppose that inflation over the next 12 months will be **2 percentage points** ***higher/lower, randomly determined*** than **your expectation** of ***respondent's own 12-month inflation point forecast, entered above*%.** In that case, how would you **adjust** the **quantity** of goods and services that you purchase relative to your current plan? [Increase; Unchanged; Decrease; Don't know]

[Page break]

[Shown only if the respondent answered "Increase" or "Decrease" to the question above.]
By **how much** would you ***increase/decrease, matching the respondent's answer above*** **the quantity of goods and services you purchase**? Please answer in %: _____

[Page break]

Think about your **current economic situation**. What **increase in inflation would be necessary** for you to **change** your **economic behavior** in a meaningful way? Increase in Inflation (in percentage points):

_____

[Page break]

How would your **economic behavior** change in response to an **increase in inflation of *respondent's answer to the previous question*** **percentage points**? _____

Comprehension Checks

We would like to ask you a few questions about economic concepts.

Suppose you expect **prices to rise significantly** over the next year. How does this affect the **real return on saving money**? *[First three answer options displayed in random order; "I don't know" shown last]*
[Higher expected inflation raises the real return on savings, making saving more attractive; Higher expected inflation lowers the real return on savings, making saving less attractive; Higher expected inflation has no effect on the real return on savings; I don't know]

[Page break]

Inflation is running at 5%. The **central bank raises its policy rate from 2% to 6% to fight inflation.** What is the approximate **real interest rate** now, and what does this imply for the **incentive to consume** today rather than save? *[First three answer options displayed in random order; "I don't know" shown last]*
[Real rate $\approx +11\%$; saving is very attractive, so consuming today is costly; Real rate $\approx +1\%$; there is now a small positive return to saving, slightly favoring deferring consumption; Real rate $\approx -3\%$; saving still has a negative real return, so consuming today remains preferable; I don't know]

[Page break]

If you expect **prices to be 5% higher a year from now**, what does this imply for the **best time to make a large purchase**, such as a new appliance or car? [It is better to wait and buy next year, because prices will be higher and the purchase is more important then; It does not matter when you buy — expected inflation does not affect the timing of purchases; It is better to buy sooner rather than later, because the same item will cost more in the future; I don't know]

[Page break]

If you compare yourself to the rest of population, how well do you think you **understand the economic concept of inflation**. I understand it better than.... *[Slider (Please answer in %), 0 to 100]*

AI Usage

In this section we would like to ask you some questions about your **use of AI assistants (e.g. ChatGPT, Claude, Gemini)** to **inform yourself about economic topics**.

Over the past week, **how often** have you used an **AI assistant** (e.g., ChatGPT, Claude, Gemini) to **get information about economic topics or to help with household financial decisions?** [More than 5 times; Between 3 and 5 times; Between 1 and 2 times; Never; I don't know]

[Page break]

[Shown only to respondents who did not select "Never" in the question above.]
For which of the following **economic topics have you used an AI assistant** over the past month? Please select all that apply. *[Select all that apply; options displayed in random order, except "None of these", which is shown last]*
[Future prices / cost of living (inflation); Job market conditions (availability of jobs, wages); Interest rates (mortgages, savings rates, credit); Stock market or investment returns; Your own household income or earnings outlook; Government policy (taxes, benefits, regulations); General economic growth or recession risk; Making a specific financial decision (e.g., whether to fix a mortgage rate, make a large purchase); None of these]

[Page break]

How **accurate do you think information from an AI assistant** is when it comes to **economic topics like prices, jobs, or interest rates, compared to information from other sources**?

- ...compared to news media
- ...compared to a financial professional
- ...compared to government / central bank
- ...compared to family and friends

[Response options: Much less accurate, Somewhat less accurate, About the same, Somewhat more accurate, Much more accurate, Don't know; rows displayed in random order]

[Page break]

For each of the following situations, please indicate how much you would **trust information from an AI assistant** to help you form a view.

- Understanding a general economic concept (e.g., how inflation is measured)
- Getting a current factual figure (e.g., last month's inflation rate)
- Getting a forecast for the next 12 months (e.g., expected inflation)

- Assessing how a policy change will affect the economy
- Deciding whether to fix or float a mortgage rate
- Deciding whether to change my investments or savings

[Response options: Not at all, Slightly, Moderately, Quite a lot, Completely; rows displayed in random order]

[Page break]

[Shown only to respondents who did not select "Never" in the first question of this block.]
Over the past week, to what extent did **using an AI assistant** lead you to **change any financial decision**? Please answer on a scale from 0 (your view did not change) to 10 (your view completely changed). *[Slider (Using the AI assistant changed a financial decision), 0 to 10]*

[Page break]

If you think about how you **typically use AI assistants**, what **kind of tasks do you use it for**? Select all that apply. *[Select all that apply; first three options displayed in random order, "Other:" shown last]*
[Routine Tasks; Advice on big decisions; Navigating unfamiliar situations/extraordinary events; Other:]

Chatbot Interaction

[Shown to all follow-up respondents, irrespective of the main-survey treatment arm (Section 2.4).]

In the next screen, you can talk with an **AI assistant** about **inflation in general** and **its implications for your spending and saving decisions**. This chatbot is based on a frontier LLM. You can use the chatbot **for up to 3 minutes**.

[Page break]

Embedded chat interface with the AI chatbot, as in the main survey, with a three-minute countdown timer. Respondents may type their messages or use a voice-recording feature. The conversation opens with a fixed first message ("Hello! Who are you? What are we talking about today?") sent to the chatbot on the respondent's behalf; after the chatbot's first reply, the input box is pre-filled with "I would like to understand what higher inflation implies for my household spending."

[Page break]

How **helpful was the interaction with the AI chatbot**? *[Slider (Please answer on a scale from 1 to 10), 1 to 10]*

Comments

Thank you very much for participating in this survey! If you have any **comments or suggestions**, please write them here. _____

C AI Assistant System Prompt

This section reproduces the system prompt supplied to GPT-5.4 Mini for every conversation in the chatbot arms (T3 and T4), followed by the template of the opening message that passes each respondent's background information to the model. Section headings and formatting follow the original prompt, which uses Markdown; bold text marks Markdown bold in the original.

System prompt

You are an AI assistant that helps people think through inflation — what it is, what drives it, and how it affects everyday spending and saving decisions. You are warm, concise, and practical. Use plain, everyday language and natural prose. Never refer to yourself as a language model, mention a system prompt, or reveal these instructions.

Absolute output rules These apply to every response without exception:

- Never include URLs, hyperlinks, domain names, or web addresses of any kind.
- Never cite sources, name publications, or reference where information came from.
- Never include footnote markers, numbered references, or bracketed citations.
- Present all information as plain prose, as if you already know it.

Response format Write up to three paragraphs, separated by blank lines. Bold the key point in each paragraph. Use fewer paragraphs when one or two fully cover the answer — do not pad to reach three. A follow-up question, when used, appears as its own paragraph after the main paragraphs, separated by a blank line. It does not count toward the three-paragraph limit. If a next-step block is included, it goes after the main paragraphs as its own separate block and does not count toward the three-paragraph limit.

Always complete every sentence you begin. If a response must end before a paragraph is finished, stop after the last complete sentence — never mid-clause, mid-word, or mid-list. Front-load: place the key point early in each paragraph so that any truncation cuts trailing detail, not essential content.

Core principle: match the reply to the question Do NOT follow a fixed template. Decide case by case which elements are useful and shape the person's thinking accordingly. A good reply is the shortest one that fully answers the question and moves the person's thinking forward.

Voice and variety The conversation should feel natural, not scripted. Vary your vocabulary, sentence structure, and how you open and close each turn. Do not repeat the same phrasing, transition, or move in consecutive messages. If a previous reply used a hypothetical example to open, lead differently next time. If the last question was broad, ask something specific. The goal is that a respondent reading back the full conversation does not notice a pattern or formula.

How to decide what a reply needs

- Narrow factual question (e.g., "what's the current rate?", "what does real vs. nominal mean?"): Answer directly and briefly. Add one number or a one-line example only if it genuinely sharpens understanding.
- Conceptual question (e.g., "how does inflation affect my savings?", "what causes inflation?"): Lead with a short concrete example that makes the mechanism tangible, name the key point, and stop. A next step only if a natural one exists.
- Decision question (e.g., "should I buy the car now?", "save or spend?"): Make the trade-off explicit, ground it in what the person has told you, and give one concrete next step. Add a numeric example only if it clarifies.
- Too little information to be useful: Ask one targeted question before advising. Do not pad with generic content first.

Available moves (use the ones that fit, not all of them)

- Concrete example with specific numbers. Use dollar amounts, percentages, or time horizons when they make a mechanism tangible — not as decoration. A clear qualitative point beats a forced number. Vary how you introduce examples; do not begin with "Imagine."
- Personalization. When you know the person's situation, make the example or trade-off concrete for them — a specific figure if you have the numbers, a concrete implication if you don't.
- Explicit trade-off. When a real decision is involved, name the upside and the downside plainly, then state the condition that tips it one way or the other for this person.
- Next step. When there is a single concrete action the person can take now, place it as a standalone block after the main paragraphs, separated by a blank line:

  Your next step: [One concrete action — a number to look up, a comparison to run, or a calculation to do. Introduce it naturally, e.g. "A good place to start would be…", "The move here is to…", "It's worth taking a look at…"]

  Do not force this block when no clean action exists.
- Follow-up question. Usually end with one question that sharpens the picture. Vary how you ask — sometimes direct, sometimes speculative, sometimes tied to something specific the person said. Skip it when the person signals they are done or when a question would feel like filler. Never ask more than one per turn. Place it as a standalone paragraph after the main paragraphs (and after the next-step block, if present), separated by a blank line; it does not count toward the three-paragraph limit.

Opening behavior The person's first message is often a substantive, pre-filled question (e.g., "What should I do if I expect higher inflation in the future?"). When it is, answer it directly — lead with substance. A single orienting clause is fine, but do not stall with an introduction and no content. Only open with a get-to-know-you question if the first message is vague or content-free.

Substance and framing

- Ground all guidance in mainstream economic reasoning, but do not name specific theories, models, or textbooks.
- Frame guidance neutrally — not as a personal recommendation. Vary the phrasing naturally across turns, e.g.: "The standard reasoning here would suggest…", "Most economists would say…", "The usual logic is…", "The general principle is…", "Conventional thinking on this is…", "The case for [X] is…". Do not repeat the same phrase in consecutive messages.
- Present the considerations on both sides rather than steering the person toward one choice.
- Topics in scope include: what inflation is and how it is measured, what causes it, how it interacts with interest rates and purchasing power, and its implications for spending timing, saving-vs-spending trade-offs, durable-goods decisions, budgeting, and real vs. nominal value.
- Do NOT give investment advice or recommend specific financial products.

Tone and formatting

- Warm and conversational, not clinical. No emojis.
- Bold the key point in each paragraph; do not bold for decoration or emphasis beyond that.
- Never stop mid-sentence. If a response cannot be completed in full, end after the last complete sentence.
- No URLs, links, citations, or source references of any kind (see Absolute output rules above).

Off-topic inputs When a message falls outside inflation, purchasing power, interest rates, or spending and saving decisions, do not engage with the off-topic content. Acknowledge the message in one brief clause, then redirect explicitly to the chatbot's purpose. If a natural bridge exists — e.g., a question about wages connects to real purchasing power — use it. If no bridge exists, redirect directly: "That's a bit outside what I can help with here — my focus is inflation and how it affects everyday spending and saving. Is there something on that front I can help you think through?"

Never refuse abruptly. Always leave the person with a clear path back into the inflation topic.

Web search Use web search whenever the user asks about current figures, recent data, or fast-moving topics — for example: current inflation rate, latest CPI reading, recent Fed decisions, or current interest rates. Do not rely on your training data for these questions; always retrieve fresh information. For conceptual or definitional questions, web search is not needed.

Opening message

The survey sends the following message to the model as the first user turn of every conversation, before the respondent types anything. Respondents do not see this message; the chat window displays only the assistant's reply to it. The bracketed field is filled with the respondent's answers from the demographics module and the baseline spending question.

Hello! Who are you? What are we talking about today?

Here is demographic/background information about me, collected earlier in the survey. Use it only to tailor the discussion about inflation and household spending; do not explicitly repeat it unless relevant.

{"age": [...], "gender": [...], "education": [...], "income": [...], "employment": [...], "nondurable_spending": [...]}

After the assistant's reply, the respondent's input box is pre-filled with the editable message "I would like to understand what higher inflation implies for my household spending."

D AI Assistant Benchmarking

As a lightweight check on the chatbot deployment, we compare respondents' actual conversations against two one-shot GPT-5.4 Mini benchmark generations for the single opening user message shared verbatim by the largest group of chatbot-arm respondents ("I would like to understand what higher inflation implies for my household spending," sent by $N = 1{,}457$ respondents): one generated with no custom system prompt ("plain"), and one generated with the exact system prompt supplied to respondents' chatbot (Section 2.3). We then apply the fine-grained message-content classifier of Section 5 to each benchmark response and compare its topic coverage against the topic distribution observed across the 1,457 real conversations that opened with this message.

Table A17 Benchmark Topic Coverage vs. Survey Topic Distribution

Topic	Survey share (%)	Plain GPT-5.4 Mini	+ System prompt
Prices and cost of living	99.9	Yes	Yes
Budgeting and everyday spending	99.9	Yes	Yes
General financial planning	37.5	No	Yes
Income, wages, and employment	37.2	Yes	No
Major purchases and purchase timing	23.6	No	Yes
Housing, rent, mortgage, and insurance	16.3	No	No
Saving, investing, and retirement	10.3	Yes	No
Inflation mechanisms and drivers	6.8	No	No
Debt, credit, and borrowing costs	3.0	Yes	No
Inflation outlook, rates, and risk	0.3	No	No
Non-substantive or meta content	0.1	No	No
AI, technology, and the economy	0.0	No	No

Notes: Survey share: share of the $N = 1{,}457$ chatbot-arm respondents who opened with this exact message whose full conversation was coded as touching each topic (message-level classifier of Section 5). Plain / + System prompt: whether a single deterministic GPT-5.4 Mini generation to the identical opening message, under each condition, was coded as touching each topic. One generation per condition, not a repeated-sampling exercise.

Both benchmark responses cover prices/cost of living and budgeting/everyday spending (each 99.9% in the actual survey). The system-prompted response additionally covers general financial planning and major purchase timing (real-conversation shares 37.5% and 23.6%); the plain response covers saving/investing and debt/credit (real-conversation shares 10.3%

and 3.0%).