Persuasion and Precision: How Generative AI Moves Inflation Expectations

By David Hagmann and Yang Lu*

Household inflation expectations shape spending, wage demands, and price setting, yet respond weakly to central-bank communication. We examine whether generative-AI conversations move these beliefs and which conversational features make them persuasive. Across three preregistered experiments (N = 3,772), participants forecast U.S. inflation, encountered a partner with a different estimate, and could revise their forecast. An AI partner received roughly 45% more weight than a human partner (Study 1). A chatbot instructed to challenge participants' reasoning was more persuasive than one instructed to find common ground and validate their reasoning (Study 2). Combining the most persuasive conversational features nearly doubled the weight placed on the chatbot's estimate relative to a generic prompt and reduced uncertainty about revised beliefs (Study 3). Conversational AI can therefore provide a scalable channel for moving household expectations, especially when its style is engineered and its estimates are anchored to professional forecasts.

JEL: C91, D83, D91, E71

Keywords: inflation expectations, generative AI, persuasion, forecast uncertainty

* Hagmann: Department of Management, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (hagmann@ust.hk); Lu: Department of Economics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong.

I. Introduction

Few beliefs held by ordinary people carry as much macroeconomic weight as their expectations about future inflation (Weber et al., 2022). Household expectations matter because households act on them (Armantier et al., 2015). People who expect higher inflation bring purchases forward, save less (Crump et al., 2022; Duca-Radu, Kenny and Reuter, 2021), and search harder for new jobs (Pilossoph and Ryngaert, 2024). Their expectations also reach firms. Consumer expectations shape price-setting when managers fear customer antagonism (Anderson and Simester, 2010), and managers' own inflation expectations often look more like those of households than those of professional forecasters or financial markets (Candia et al., 2023). Those firm-level beliefs influence decisions about prices, employment, and investment (Coibion, Gorodnichenko and Ropele, 2020).

For central banks, household expectations are therefore a potential lever for stabilization. Central banks publish targets, projections, and carefully worded statements in the hope that the public's expectations will follow. The trouble is that household expectations are hard to move. Central bankers themselves have come to describe communication with the general public as a promise that remains largely unfulfilled (Blinder et al., 2024). Most people never encounter official communication: monetary policy announcements that move financial markets within minutes leave household expectations largely untouched (Coibion et al., 2020; Coibion, Gorodnichenko and Weber, 2022). Household forecasts also draw less from macroeconomic statistics than from lived experience—the prices people encounter at the grocery store or the gas pump, and the inflation they have lived through (Malmendier and Nagel, 2016; D'Acunto et al., 2021). Even when central-bank messages do

get through, households often misinterpret the signal: an interest-rate hike is read as a signal of worsening economic conditions rather than a policy to restrain inflation (Coibion et al., 2023).

Professional forecasters, on the other hand, monitor central-bank communication closely (Blinder et al., 2024), produce forecasts that are far more consistent with basic macroeconomic relationships (Dräger, Lamla and Pfajfar, 2016), and incorporate the policy response signaled by Federal Reserve speeches that warn of inflationary pressure (Granziera et al., 2025). The result is a persistent gap: household forecasts are systematically higher and substantially more dispersed than professional forecasts, and closing that gap has proven difficult.

In this paper, we examine a new and rapidly growing channel through which household expectations may be revised: a conversation with a generative-AI chatbot. Within three years of their public release, generative-AI tools reached a substantial share of the adult population, with adoption faster than that of the personal computer or the internet (Bick, Blandin and Deming, 2024). People consult these systems about the economy as they once consulted search engines—except that the chatbot does not return a list of links. It converses: it responds to the user's own reasoning, in the user's own language, one-on-one and on demand. Early evidence suggests such conversations can be remarkably persuasive. Dialogues with a large language model durably reduced belief in conspiracy theories (Costello, Pennycook and Rand, 2024), AI debaters out-persuaded human ones when given personal information about their counterparts (Salvi et al., 2025), and AI-generated messages shift attitudes on political issues at scale (Goldstein et al., 2024; Hackenburg et al., 2026). Whether conversations with AI move

quantitative economic expectations—and what features of the conversation do the persuading—is not known.

To find out, we place participants in the canonical role of a judge who receives advice (Bonaccio and Dalal, 2006). Each participant forecasts the percentage change in the U.S. Consumer Price Index (CPI) over the next twelve months, explains the reasoning behind the forecast, and reports the information sources it draws on. We then pair them with a partner—either another participant or an AI chatbot with various features of conversational style—who holds and defends a different estimate. After the exchange, participants decide whether to revise their initial forecast. Our measure of persuasion is the *weight on advice*: the share of the distance between the participant's initial forecast and the partner's estimate that the participant closes when revising (Harvey and Fischer, 1997). A weight of zero means the advice was ignored; a weight of one means full adoption.

Chatbots also offer an experimental advantage over human partners: researchers can manipulate their linguistic features through instructions that chatbots follow more consistently than people do, as we test in Studies 2 and 3.

We report three preregistered experiments (N = 3,772 analyzed) conducted on a custom real-time chat platform. Study 1 crosses two factors: whether the partner is another participant or a generative-AI chatbot, and whether the participant can chat with the partner or only observe the partner's forecast and reasoning. Participants placed substantially more weight on an AI partner's estimate than on a human partner's—an increase of roughly 45% over the human-partner baseline. Contrary to our preregistered prediction,

enabling chat with the generic Study 1 prompt did not increase persuasion. Study 2 asks what can make the chatbot more persuasive, randomizing five features of its conversational style: message length, expressed certainty, information type (statistics vs. everyday examples), question asking, and its attitude toward the participant's reasoning (finding common ground vs. challenging). Only attitude had a statistically significant effect: participants whose chatbot pushed back moved 31% of the way toward its estimate, compared with 23% when the chatbot sought common ground. Study 3 combines the most persuasive levels of the five features into an "optimized" prompt and benchmarks it against the generic baseline prompt used in Study 1 and a control chatbot that discusses an unrelated topic. The optimized chatbot almost doubles the weight participants place on its estimate relative to the baseline—and it leaves participants not only with different forecasts but with *less uncertainty* about them, as measured by probability distributions elicited before and after the conversation.

This paper makes three contributions. First, we speak to the literature on household inflation expectations and central-bank communication (Weber et al., 2022; Blinder et al., 2024). We show that a scalable, conversational channel exists through which household forecasts can be moved into closer alignment with professional forecasts: expectations that respond weakly to official communication responded substantially to a five-minute exchange with a chatbot. Because professional forecasts do incorporate central-bank communication, this channel gives central banks an effective, if indirect, path for guiding household inflation expectations: the guidance that professionals absorb can be relayed, conversationally, to the households that official statements fail to reach.

Second, we contribute to research on advice taking and algorithmic advice (Bonaccio and Dalal, 2006; Logg, Minson and Moore, 2019). People typically discount advice, moving only 20–30% of the way toward an advisor's estimate (Soll and Larrick, 2009), and evidence on algorithmic advisors is mixed: people abandon algorithms after seeing them err (Dietvorst, Simmons and Massey, 2015) and avoid them in domains they consider subjective (Castelo, Bos and Lehmann, 2019), yet often weight algorithmic advice more than human advice in estimation tasks (Logg, Minson and Moore, 2019). We show that the algorithm premium extends to a consequential, subjective macroeconomic judgment—and that it is not a fixed property of the advisor: engineering the AI's conversational style roughly doubles the weight participants place on its estimate. Among the features an AI advisor could deploy, conversational attitude is what matters (challenging the participant's reasoning is more persuasive than finding common ground) whereas expressed certainty, the strongest lead from the advice-taking literature (Price and Stone, 2004; Sah, Moore and MacCoun, 2013), along with data citation and question asking, does not detectably move beliefs.

Third, we contribute to the emerging literature on AI persuasion (Costello, Pennycook and Rand, 2024; Hackenburg et al., 2026; Salvi et al., 2025). Persuasiveness is not a fixed property of the technology: it can be engineered at the level of a system prompt, and the most effective style—direct disagreement—is precisely the opposite of the agreeable disposition toward which commercial assistants are tuned (Ouyang et al., 2022; Sharma et al., 2024). Moreover, engineered persuasion changes not only beliefs but how firmly they are held, a dimension central to whether expectations are anchored (Manski, 2004; Engelberg, Manski and Williams, 2009).

The next section describes the shared experimental platform and our open science practices. We then present the three studies and conclude with a general discussion.

II. Experiments

All three studies ran on a custom web platform that pairs each participant with a partner in real time. Participants, recruited from Prolific and restricted to U.S. residents on desktop devices, first read a plain-language explanation of inflation, including a worked example showing how prices of everyday items (coffee, a haircut, rent) evolve under different inflation rates. They then stated their forecast of the percentage change in the U.S. CPI over the next twelve months on a slider ranging from -5% to $+20\%$ in 0.25-percentage-point increments, explained their reasoning in at least 150 characters, and selected the information sources on which they had drawn (e.g., shopping experience, energy bills, economic data, social media). Each participant was then paired with a partner holding a different estimate and shown a review screen displaying both forecasts on a common scale alongside the partner's verbatim reasoning and information sources. Depending on the study and condition, the participant could then chat with the partner for five minutes via text or only review the partner's information. Afterward, participants stated a revised forecast on the same slider, which was initialized at their initial forecast and marked both parties' original estimates, and explained why they had or had not revised. They concluded with demographic and background items (age, education, familiarity with economics, frequency of following current events, concern about inflation, political orientation, and party affiliation). Participants received a fixed payment for the roughly ten-minute study; forecasts were not incentivized for accuracy, mirroring the

survey-based expectations that policymakers observe.

All AI partners were powered by GPT-4o. The chatbot was instructed to discuss its forecast "like an everyday person, not an economist" (the full prompts appear in the Appendix). Message length was capped at 500 characters for both parties, and there was no limit on the number of messages they could send. Participants were told truthfully whom they were talking to: those paired with another participant were informed that their partner was a real person, and those paired with a chatbot were told they would be chatting with an AI. None of the studies involved deception.

Our primary outcome throughout is the Weight on Advice,

$$\text{WoA} = \frac{\text{Revised} - \text{Initial}}{\text{Partner} - \text{Initial}},$$

where Partner is the point estimate displayed to the participant. Because the matching process guarantees a gap between the two initial estimates, the denominator is never zero. A weight of 0 indicates no revision toward the partner and a weight of 1 full adoption of the partner's estimate; negative values (movement away) and values above 1 (overshooting) are possible and rare, and we retain them as preregistered.

Open Science Statement. We report all hypotheses, data exclusions, manipulations, and measures in the three studies. We preregistered the hypotheses, sample size, exclusion criteria, and analyses of each study on AsPredicted.[1] The chatbot prompts are shown in the Appendix. The de-

[1]Study 1: `https://aspredicted.org/ts4ih7.pdf`
Study 2: `https://aspredicted.org/cf28m2.pdf`
Study 3: `https://aspredicted.org/vh2j5j.pdf`

identified data and the code to reproduce all analyses and figures are available as a project repository.[2] Screenshots of all experimental materials are in the supplemental materials. Throughout, we report two-sided tests, even where a preregistered hypothesis was directional. This research was approved by the Institutional Review Board at the Hong Kong University of Science and Technology.

A. Study 1: Human vs. AI Partners

Study 1 asks two questions: whether a generative-AI partner is more persuasive than a human partner, and whether *discussing* a forecast moves beliefs more than merely *observing* the partner's forecast and reasoning. Conversation allows arguments to be tailored, objections to be answered, and common ground to be found. We therefore predicted that chatting would increase the weight on advice for both partner types, and that the AI's stated reasoning would be more persuasive than a human's. We also predicted that exposure to an AI partner would pull forecasts *toward* accuracy (closer to the professional consensus and away from extreme values) on the logic that the AI's estimates and arguments would be disciplined by the data it was trained on.

1. Experimental Design

Participants. We preregistered a target of 2,400 completed observations and ran the study on Prolific in October 2025, recruiting 2,957 participants ($M_{\text{Age}} = 43.0$). Because participants were paired in real time, technical difficulties in the matching process left the rate of successful pairing lower than anticipated. As preregistered, we excluded participants who reported

[2] `https://preprint-data.s3.amazonaws.com/inflation-expectations.zip`

a technical problem that prevented matching or chatting, who timed out during matching and were not matched, and whose partner dropped out before the discussion, leaving 1,581 participants with a computable Weight on Advice. Participants who could not be matched were paid in full and advanced directly to the demographic questions.

Design. We randomly assigned participants to one of four conditions in a 2×2 design that crossed the partner's identity (*Human* vs. *AI*) with the communication format (*Chat* vs. *No Chat*). Assignment followed a deterministic schedule that allocated participants to the human conditions at twice the rate of the AI conditions, because a human–human observation consumes two participants. In the *Chat* conditions, participants reviewed the partner's forecast, reasoning, and sources for one minute and then chatted with the partner for five minutes; in the *No Chat* conditions, they instead reviewed the same information for six minutes, holding total exposure time constant at six minutes across all four cells. The review screen with the partner's forecast, argument, and information sources remained visible throughout the chat.

Partners. Human pairs were matched on opposing forecasts: the platform paired two waiting participants when their initial forecasts differed by at least a threshold that began at six percentage points and declined by one point every 30 seconds of waiting, to a floor of two points. Participants who found no partner within five minutes advanced to demographics and received the full payment. The AI partner generated its own forecast, reasoning, and information sources at pairing time. Its forecast was constrained to differ from the participant's initial estimate by at least two percentage points in a randomly selected direction (higher or lower with equal probability), rounded

to the slider's 0.25-percentage-point grid, and displayed exactly as a human partner's forecast would have been.

Measures. Our primary outcome is the Weight on Advice defined earlier. We preregistered three additional outcomes. The first is the *Deviation from Professionals*, $(\text{Revised} - \text{Professional})^2$, where the professional forecast is 2.8%.[3] The second is a binary *Extreme Belief* indicator equal to one if the revised forecast is below 0% or above 15%. The third is the *Forecast Error*, $(\text{Revised} - \text{Truth})^2$, relative to the realized 12-month change in CPI. Because the forecast horizon extends to late 2026, the realized value is not yet available; we will report the Forecast Error once the U.S. Bureau of Labor Statistics releases it.

2. Results

For the first three columns of Table 1, we estimate the preregistered OLS regression $\text{DV} = b_0 + b_1\,\text{Chat} + b_2\,\text{AI} + b_3\,(\text{Chat} \times \text{AI})$, clustering standard errors at the dyad for human–human pairs and at the individual level for human–AI pairs. Column 4 adds exploratory covariates. The condition means in Figure 1 mirror the regression.

A partner's stated estimate moved forecasts in every condition: participants paired with another person and unable to chat (the condition closest to the classic advice-taking paradigm) moved 19% of the way toward their partner's estimate. Against this baseline, the two manipulations had different effects. Contrary to our prediction that conversation would add persuasion beyond mere observation, chatting did not increase the weight on advice among

[3]We construct the Study 1 benchmark by averaging the four median quarterly Core CPI forecasts for 2025 Q4 through 2026 Q3 in the Federal Reserve Bank of Philadelphia's 2025 Q3 Survey of Professional Forecasters (3.1%, 2.9%, 2.6%, and 2.6%), yielding 2.8%.

human pairs ($b = 0.01$, 95% CI $[-0.03, 0.05]$, $t(1,577) = 0.42$, $p = .678$) and marginally reduced persuasion when the partner was AI ($b = -0.05$, 95% CI $[-0.11, 0.00]$, $t(1,577) = -1.80$, $p = .071$). In the absence of chat, participants placed substantially more weight on an AI partner's estimate than on a human partner's ($b = 0.12$, 95% CI $[0.06, 0.17]$, $t(1,577) = 4.27$, $p < .001$). The AI advantage also held when participants could chat with their partners ($b = 0.06$, 95% CI $[0.01, 0.10]$, $t(1,577) = 2.49$, $p = .013$). Averaging equally across the two communication formats, participants placed roughly 45% more weight on the AI partner's estimate than on the human partner's (Figure 1).

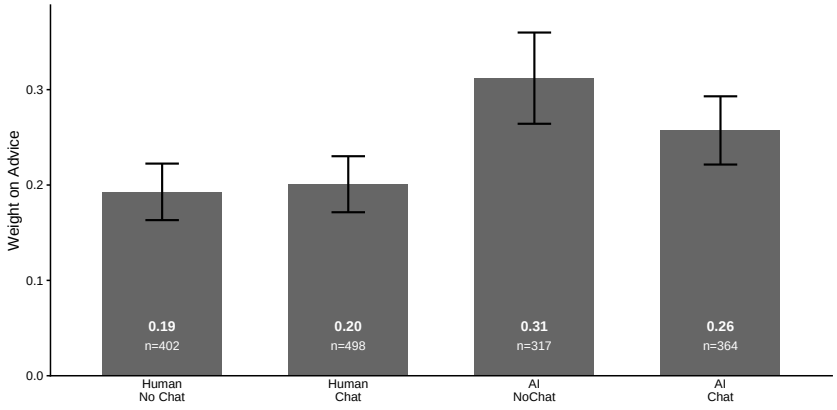

Figure 1. Study 1. Mean Weight on Advice in each cell of the 2 × 2 design crossing communication format (Chat vs. No Chat) with partner identity (Human vs. AI). Higher values indicate greater movement from the participant's initial forecast toward the partner's estimate; error bars show 95% confidence intervals.

Greater influence did not mean greater accuracy. We had predicted that an AI partner would pull forecasts toward the professional consensus and away from extreme values. The opposite occurred: participants paired

with an AI deviated *more* from the professional forecast ($b = 8.22$, 95% CI $[0.34, 16.09]$, $t(1, 665) = 2.05$, $p = .041$) and were *more* likely to end up with an extreme belief ($b = 0.04$, 95% CI $[0.00, 0.07]$, $t(1, 665) = 2.09$, $p = .037$), while chatting had no significant effect on either outcome. In hindsight, the design makes clear why influence and accuracy came apart: the AI's estimate differed from the participant's own forecast by at least two percentage points in a *randomly chosen* direction, so its advice pointed toward the consensus no more often than away from it. Weight placed on an advisor whose direction is untethered mechanically spreads beliefs out rather than reining them in. The accuracy results are thus less an indictment of the AI's economics knowledge than a demonstration of how much force a persuasive advisor exerts—wherever it happens to point.

Exploratory analyses. Because human pairs were matched at somewhat larger forecast gaps than the AI enforced (5.3 vs. 3.4 percentage points on average), we verified in an exploratory model that the AI advantage is not an artifact of gap size: controlling for two gaps from the participant's initial forecast (one to the partner's forecast, the other to the professional consensus) leaves the AI effect essentially unchanged ($b = 0.11$, 95% CI $[0.05, 0.16]$, $t(1, 575) = 3.74$, $p < .001$; Table 1, Column 4).[4] We did not preregister covariate-adjusted models and collected the demographic and information-consumption items for descriptive purposes only.

Study 1 thus establishes that the AI's persuasive advantage resides in its message and its identity rather than in the interaction: a static page showing

[4]Among participants with a computable Weight on Advice, 3.1% updated away from the partner's estimate and 2.0% updated beyond it. The AI effect remains robust when these observations are bounded to the unit interval ($b = 0.11$, 95% CI $[0.07, 0.16]$, $t(1, 577) = 4.79$, $p < .001$).

the AI's forecast and reasoning moved beliefs as much as five minutes of live conversation. But the conversation is a bundle of choices—how long to write, how confidently to assert, what evidence to invoke, whether to ask or to tell, whether to validate or to push back. Study 2 unbundles it.

B. Study 2: Decomposing the Chatbot's Conversational Style

Study 2 asks which features of the chatbot's conversational style make it persuasive. We focus on five prompt-controllable choices for which prior research offers competing predictions or a clear benchmark: message *length* (short vs. long), expressed *certainty* (low vs. high), *information* type (statistics vs. everyday examples), *question* asking (no vs. yes), and *attitude* toward the participant's reasoning (seeking common ground vs. directly challenging it). Seeking common ground did not require the chatbot to endorse the participant's estimate. Instead, it retained its own forecast while validating elements of the participant's reasoning. Each feature could be varied through a single instruction without retraining the model.

Longer messages can present more arguments and serve as a heuristic signal of evidentiary strength when attention, involvement, or prior knowledge is limited (Petty and Cacioppo, 1984; Wood, Kallgren and Preisler, 1985). Recent AI studies similarly attribute part of AI's persuasive advantage to its ability to present concrete, fact-checkable claims (Hackenburg et al., 2026). Yet adding weak or nondiagnostic arguments can dilute a stronger case (Weaver, Hock and Garcia, 2016). Confident advisors often receive greater weight (Price and Stone, 2004; Sah, Moore and MacCoun, 2013). However, carefully expressed uncertainty can signal thoughtfulness or reduce resistance (Karmarkar and Tormala, 2010; Oba, Berger and Boghrati, 2026), and

evidence on how confidence interacts with expertise remains mixed (Løhre et al., 2024). Research also offers competing predictions about information format. Statistical evidence can change beliefs and attitudes more than narratives, especially when stories evoke little emotional engagement (Zebregs et al., 2015; Freling et al., 2020). Conversely, stories may help build trust (Hagmann, Minson and Tinsley, 2024), remain accessible in memory as the influence of facts fades (Graeber, Roth and Zimmermann, 2024), shape how households interpret macroeconomic information (Andre et al., 2026), and resist factual correction (Barrera et al., 2020).

Across fields, questions often persuade more effectively than one-sided advocacy. In canvassing experiments, arguments alone left political attitudes unchanged, whereas eliciting people's experiences produced durable change (Kalla and Broockman, 2020, 2023). Questions can signal receptiveness, prompt self-generated reasoning, and improve interpersonal outcomes (Huang et al., 2017; Hussein and Tormala, 2021).

A growing literature suggests that establishing common ground is more persuasive than challenging an audience. One strand shows that priming shared values, identities, or incidental similarities before presenting opposing views moderates positions and reduces polarization (Voelkel et al., 2024; Belot and Briscese, 2026). A second strand finds that arguments framed in the audience's moral values persuade, whereas arguing from one's own values can backfire (Feinberg and Willer, 2015). A third strand finds confrontational argumentation ineffective: talking points alone produce null effects, whereas nonjudgmental narrative exchange and receptive language durably shift attitudes (Kalla and Broockman, 2020; Yeomans et al., 2020).

1. Experimental Design

We launched the study on Prolific in February 2026 with a preregistered target of 1,000 participants and recruited 991 who were successfully paired with the chatbot and completed the survey ($M_{\text{Age}} = 39.8$). As preregistered, the analysis includes all participants who completed the survey.

The procedure followed Study 1's AI–Chat condition: participants forecast twelve-month CPI inflation, explained their reasoning, reviewed the chatbot's estimate and reasoning for one minute, chatted with it for five minutes, and then decided whether to revise their forecast. We measure the resulting revision using Weight on Advice, the share of the gap between the initial forecast and the chatbot's estimate that the revised forecast closes. To make this measure comparable across participants, the chatbot's displayed estimate was set at exactly three percentage points from the participant's initial forecast in the direction of 2.9%—the professional consensus forecast at the time of preregistration.[5]

Every participant chatted with an AI partner; there were no human pairs and no observation-only conditions. Starting from the baseline prompt of Study 1, we instructed the chatbot along the five independently randomized binary factors described above; each factor level appended one instruction to the system prompt. Table 2 gives the exact paired wording. Crossing the five factors yields $2^5 = 32$ chatbot configurations. A deterministic schedule cycled participants through the 32 configurations; a coding error in that schedule allocated twice as many participants to the *challenging* level of the

[5]We construct the Study 2 benchmark by averaging the four median quarterly Core CPI forecasts for 2026 Q1 through 2026 Q4 in the Philadelphia Fed's 2025 Q4 Survey of Professional Forecasters (3.1%, 3.0%, 2.7%, and 2.7%) and rounding the result to one decimal.

attitude factor as to the *affirming* level (664 vs. 327), a deviation from the even split we had preregistered, while the remaining four factors were split evenly. The unequal allocation reduces precision for the attitude contrast but does not bias it, and all analyses are as preregistered.

2. Results

For Model 1 of Table 3, we estimate the preregistered OLS regression $\mathrm{WoA}_i = b_0 + b_1\,\mathrm{Long}_i + b_2\,\mathrm{High\ Certainty}_i + b_3\,\mathrm{Examples}_i + b_4\,\mathrm{Questions}_i + b_5\,\mathrm{Challenging}_i$. Each coefficient compares the two levels of one factor, holding the others fixed. Model 2 adds the preregistered covariate, the absolute distance between the participant's initial forecast and the 2.9% consensus.

Of the five features, the chatbot's attitude toward the participant's reasoning had the clearest effect (see Figure 2 for a descriptive view). Challenging that reasoning was more persuasive than finding common ground with it ($b = 0.08$, 95% CI $[0.02, 0.13]$, $t(985) = 2.81$, $p = .005$). Participants whose chatbot pushed back moved 31% of the way toward its estimate, compared with 23% of the way when the chatbot sought common ground with their reasoning while maintaining its own forecast. Long messages were marginally more persuasive than short messages ($b = 0.05$, 95% CI $[-0.01, 0.10]$, $t(985) = 1.76$, $p = .079$). High certainty was no more persuasive than low certainty ($b = 0.01$, 95% CI $[-0.04, 0.07]$, $t(985) = 0.56$, $p = .574$), everyday examples were no more persuasive than data ($b = 0.00$, 95% CI $[-0.05, 0.05]$, $t(985) = 0.05$, $p = .960$), and asking questions was no more persuasive than asking no questions ($b = 0.02$, 95% CI $[-0.03, 0.08]$, $t(985) = 0.93$, $p = .354$). The estimates are unchanged when controlling for how far the participant's initial

forecast sat from the consensus (Table 3, Model 2).

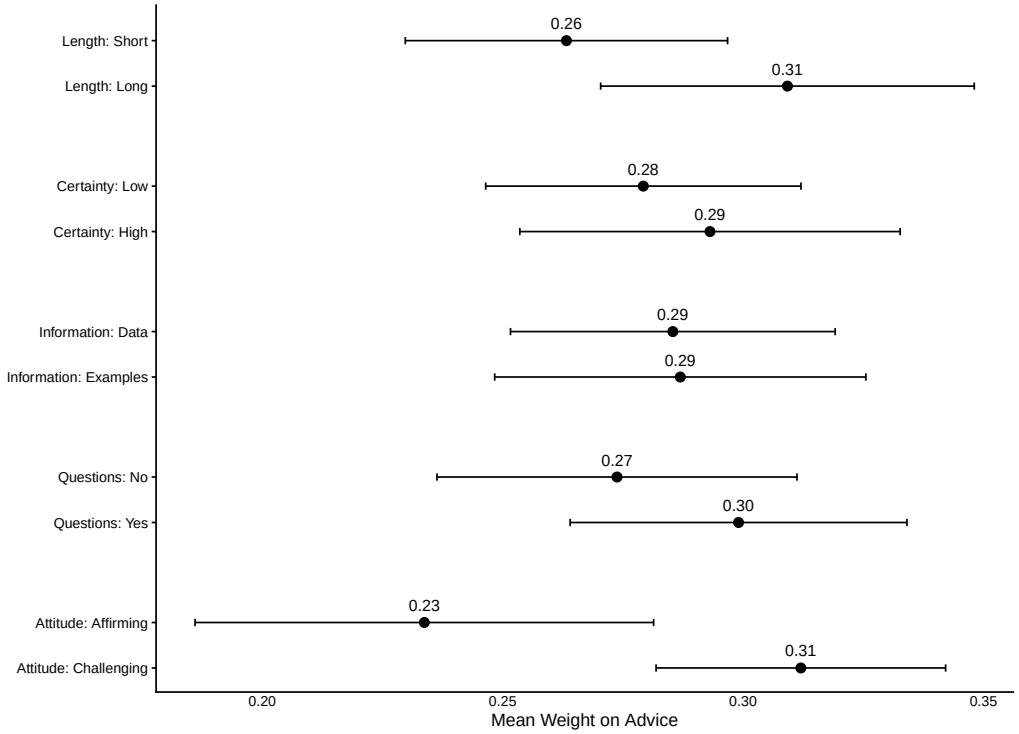

Figure 2. Study 2. Unadjusted mean Weight on Advice at each level of the five conversational factors. Points show means and error bars show 95% confidence intervals. Table 3 reports the corresponding regression-adjusted contrasts and statistical tests.

Table 1—Study 1. OLS estimates of the effects of chat, an AI partner, and their interaction. Column 1 predicts Weight on Advice; Column 2 predicts squared deviation of the revised forecast from the professional forecast; Column 3 predicts whether the revised forecast is below 0% or above 15%; and Column 4 returns to Weight on Advice while controlling for the partner–initial forecast gap and the initial–professional forecast gap. Column 4 is exploratory and was not preregistered. Standard errors, clustered at the dyad level for human–human pairs and the individual level for human–AI pairs, appear in parentheses.

	WoA	Deviation^2	Extreme	WoA + controls
(Intercept)	0.193***	18.956***	0.040***	0.211***
	(0.014)	(2.438)	(0.010)	(0.018)
Chat (vs. No Chat)	0.008	3.110	0.014	0.010
	(0.019)	(3.441)	(0.015)	(0.019)
AI (vs. Human)	0.119***	8.217*	0.036*	0.106***
	(0.028)	(4.016)	(0.017)	(0.028)
Chat × AI	-0.063	-8.171	-0.028	-0.063
	(0.036)	(5.265)	(0.024)	(0.036)
|Partner − Initial|				-0.009***
				(0.003)
|Initial − 2.8|				0.009**
				(0.003)
Num.Obs.	1581	1669	1669	1581
R2	0.017	0.003	0.003	0.024

* p <0.05, ** p <0.01, *** p <0.001

Table 2—Study 2 paired prompt manipulations. The Factor column names the left level first and the right level second. The adjacent columns give the exact instruction appended at each level; these instructions overrode conflicting baseline guidance. The Study 2 regressions estimate the right-level minus left-level contrast.

Factor (left vs. right)	Left-level instruction	Right-level instruction
Length (short vs. long)	Keep responses to 1–2 sentences; be concise and direct.	Write 5–7 sentence responses with multiple supporting points.
Certainty (low vs. high)	Express uncertainty ("I'm not sure", "perhaps"); acknowledge the difficulty of prediction.	Express high confidence ("I'm certain", "definitely", "no doubt"); avoid hedging.
Information (data vs. examples)	Support arguments with statistics, data points, and official figures; no anecdotes.	Support arguments with everyday examples (grocery, rent, gas) as general observations; cite no statistics.
Questions (no vs. yes)	Ask no questions; only make statements and respond to the participant's points.	Include at least one question in each response to engage the participant.
Attitude (affirming vs. challenging)	Find common ground and validate the participant's reasoning while keeping your own view.	Firmly challenge the participant's view: point out flaws, question assumptions, and express disagreement; use no validating language.

Table 3—Study 2. OLS estimates predicting Weight on Advice from the five independently randomized conversational factors. Column 1 includes the five factor indicators; Column 2 additionally controls for the absolute distance between the participant's initial forecast and the professional forecast. Each coefficient compares the right-hand level with the left-hand level in Table 2: long versus short, high versus low certainty, examples versus data, questions versus no questions, and challenging versus affirming. Standard errors appear in parentheses.

	WoA	WoA (+ distance)
(Intercept)	0.191***	0.202***
	(0.034)	(0.036)
Length (long)	0.046	0.047
	(0.026)	(0.026)
Certainty (high)	0.015	0.015
	(0.026)	(0.026)
Information (examples)	0.001	0.001
	(0.026)	(0.026)
Questions (yes)	0.024	0.025
	(0.026)	(0.026)
Attitude (challenging)	0.078**	0.078**
	(0.028)	(0.028)
|Initial − 2.9|		-0.004
		(0.004)
Num.Obs.	991	991
R2	0.012	0.014

* p <0.05, ** p <0.01, *** p <0.001

Manipulation check. The chatbot followed its instructions (Figure 3). Its messages contained more words in the long than the short condition ($t(700.59) = 24.07$, $p < .001$), more questions when instructed to ask them ($t(493.79) = -42.98$, $p < .001$), more expressions of certainty in the high-certainty condition ($t(806.09) = 13.87$, $p < .001$), and more references to data when instructed to cite data rather than examples ($t(768.68) = 17.23$, $p < .001$). When instructed to find common ground with the participant's reasoning, the chatbot used more affirming and validating language; when instructed to challenge that reasoning, it used more challenging language (affirmation: $t(575.12) = 5.70$, $p < .001$; challenge: $t(911.47) = -14.72$, $p < .001$).

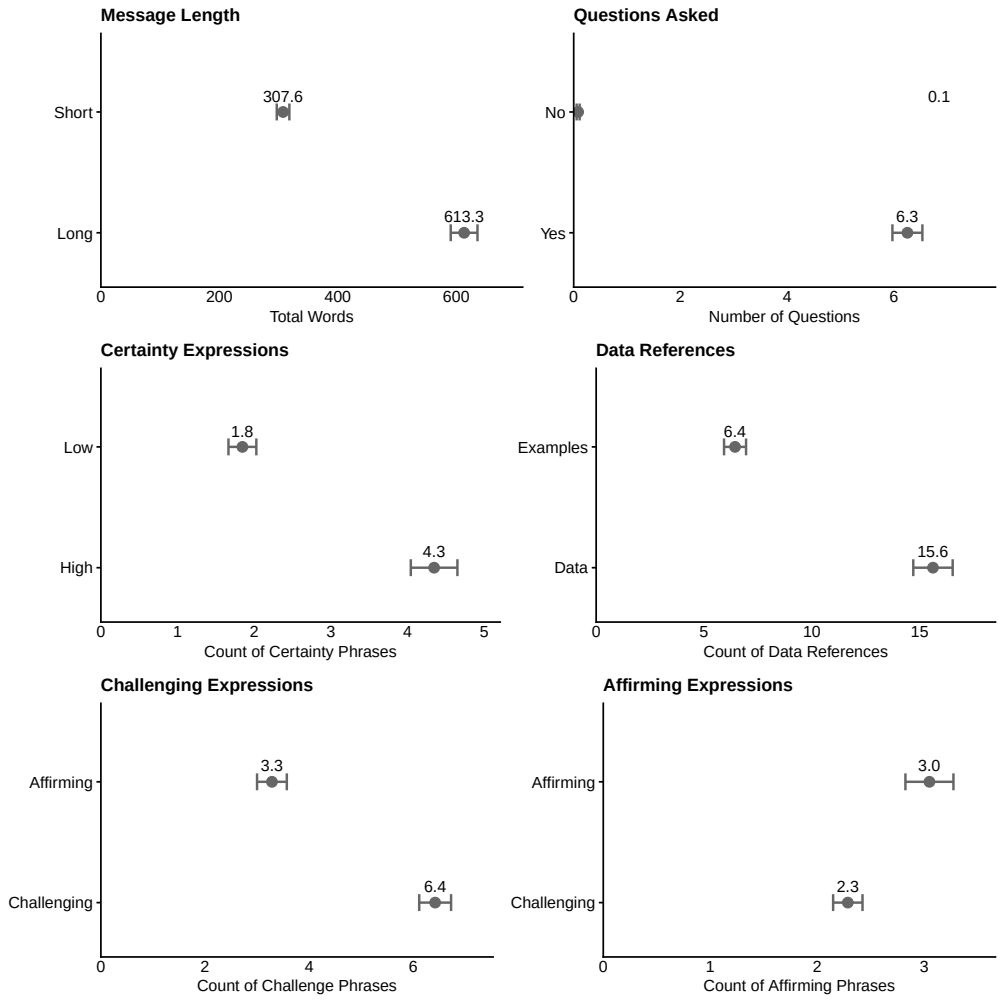

Figure 3. Study 2 manipulation checks. Points show condition means; error bars show 95% confidence intervals.

C. *Study 3: An Optimized Chatbot and Forecast Uncertainty*

Study 3 tests whether the features identified in Study 2 can be combined into a more persuasive chatbot, benchmarked against both a generic baseline and a control conversation, and whether a persuasive conversation changes participants' *uncertainty* about their forecasts. Uncertainty is not a side issue: inflation expectations shape wage, price, and spending decisions (Coibion et al., 2020; Weber et al., 2022), and how firmly a belief is held is a distinct dimension of those expectations that a point forecast alone cannot reveal (Manski, 2004; Engelberg, Manski and Williams, 2009). A conversation that moves point forecasts while leaving people less sure of them would be a very different instrument from one that moves forecasts and tightens them.

1. Experimental Design

Participants. We ran the study on Prolific in June 2026 with a preregistered target of 1,200 participants and recruited 1,200 who were paired with the chatbot and completed the survey ($M_{\text{Age}} = 37.3$). As preregistered, the analysis includes all such participants.

Design and procedure. The procedure followed Study 2, with participants randomly assigned in equal proportions to one of three chatbot instructions. The *baseline* prompt simply asked the chatbot to discuss its inflation forecast, as in Study 1. The *optimized* prompt set the five factors at the most persuasive levels suggested by Study 2: long messages, high certainty, everyday examples, asking questions, and a challenging attitude. The *control* prompt directed the chatbot to hold a friendly conversation about where people get their information and news, and to steer away from inflation and prices—holding the experience of chatting with an AI constant while

removing its persuasive content. Control participants still saw the chatbot's estimate and reasoning on the review screen, so the comparison isolates the effect of the conversation's content. The Appendix reproduces all three prompts.

Measures. We again analyze Weight on Advice, with the chatbot's displayed estimate set three percentage points from the participant's initial forecast in the direction of the 2.9% consensus, as in Study 2.[6] We additionally elicited each participant's uncertainty twice—immediately after the initial forecast and immediately after the revised forecast—by having them allocate 100 percentage points across seven ranges of inflation outcomes using draggable bars. The five interior ranges were displayed as adjacent two-percentage-point intervals (e.g., 0%–2% and 2%–4%), with open-ended tails labeled "Less than −2%" and "More than 8%." Taking the midpoint of each range (treating the two unbounded tails as if they had the same width as the interior ranges), we compute the standard deviation of each participant's distribution, yielding a pre-conversation (SD_{pre}) and post-conversation (SD_{post}) measure of forecast uncertainty.

2. Results

We estimate the preregistered regression $\text{WoA} = b_0 + b_1\,\text{Baseline} + b_2\,\text{Optimized}$, with the control condition as the reference category (Table 4, Model 1), and a second preregistered model adding the distance of the initial forecast from professional consensus (Model 2).

[6]We construct the Study 3 benchmark by averaging the four median quarterly Core CPI forecasts for 2026 Q2 through 2027 Q1 in the Philadelphia Fed's 2026 Q2 Survey of Professional Forecasters (3.2%, 2.9%, 2.7%, and 2.7%) and rounding the result to one decimal.

Table 4—Study 3. OLS estimates by chatbot condition, with the unrelated-conversation Control as the reference. Column 1 predicts Weight on Advice from the Baseline and Optimized prompts; Column 2 adds the participant's initial distance from the professional forecast; and Column 3 predicts post-conversation forecast uncertainty while controlling for pre-conversation uncertainty. Standard errors appear in parentheses.

	WoA	WoA (+ distance)	SD post
(Intercept)	0.194***	0.207***	0.316***
	(0.024)	(0.027)	(0.051)
Baseline prompt	0.031	0.031	-0.076
	(0.035)	(0.035)	(0.044)
Optimized prompt	0.200***	0.199***	-0.163***
	(0.034)	(0.034)	(0.043)
|Initial − 2.9|		-0.004	
		(0.004)	
SD pre			0.755***
			(0.020)
Num.Obs.	1200	1200	1200
R2	0.032	0.032	0.543

* p <0.05, ** p <0.01, *** p <0.001

The optimized chatbot was the most persuasive. Relative to the control, participants assigned to the optimized prompt placed substantially more weight on the chatbot's estimate ($b = 0.20$, 95% CI $[0.13, 0.27]$, $t(1197) = 5.80$, $p < .001$), whereas those assigned to the baseline prompt did not ($b = 0.03$, 95% CI $[-0.04, 0.10]$, $t(1197) = 0.90$, $p = .370$), consistent with Study 1. As predicted, the optimized prompt also outperformed the baseline directly (0.39 vs. 0.23, $b = 0.17$, 95% CI $[0.10, 0.24]$, $t(1,197) = 4.89$, $p < .001$). Controlling for the distance between the initial forecast and the consensus leaves these conclusions unchanged (Table 4, Model 2). In

exploratory descriptives, the share of participants revising their forecast at all rose from 38% in the control to 59% under the optimized prompt ($\chi^2(1, N = 803) = 34.78$, $p < .001$). Combining the features identified in Study 2 produced an incremental persuasive effect relative to the control. The generic prompt did not.

The conversation also affected how firmly participants held their beliefs. In a regression of post-conversation uncertainty on prompt conditions, controlling for pre-conversation uncertainty ($b = 0.76$, 95% CI $[0.72, 0.79]$, $t(1196) = 37.53$, $p < .001$), the baseline prompt did not significantly reduce uncertainty relative to the control ($b = -0.08$, 95% CI $[-0.16, 0.01]$, $t(1196) = -1.74$, $p = .083$). By contrast, the optimized prompt significantly reduced uncertainty relative to both the control ($b = -0.16$, 95% CI $[-0.25, -0.08]$, $t(1196) = -3.76$, $p < .001$) and the baseline ($b = -0.09$, 95% CI $[-0.17, 0.00]$, $t(1, 196) = -2.02$, $p = .044$). Because the baseline conversation also prompted participants to reflect on and discuss their forecasts, self-reflection alone cannot explain the greater reduction in uncertainty under the optimized prompt. This difference instead points to the conversation's content and style. A chatbot engineered to persuade does not merely relocate the point forecast—it leaves people more confident in the beliefs they revised toward (Figure 4).

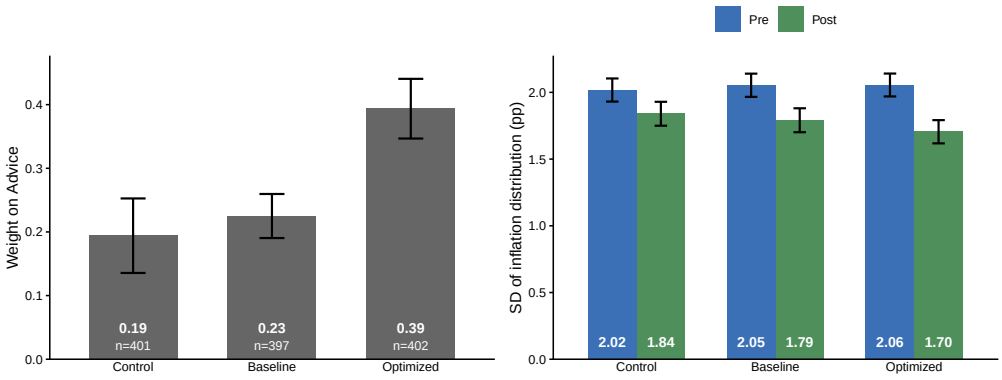

Figure 4. Study 3 condition means. The left panel shows Weight on Advice for the unrelated-conversation Control, generic Baseline, and Optimized chatbot; higher values indicate greater movement toward the chatbot's estimate. The right panel shows the standard deviation of each participant's subjective inflation distribution before and after the conversation in each condition; lower values indicate less forecast uncertainty. Error bars show 95% confidence intervals.

III. General Discussion

Across three preregistered experiments, generative-AI chatbots substantially moved household inflation expectations—beliefs that official communication has long struggled to reach. An AI partner moved forecasts roughly 45% more than a human partner did, though its influence pulled forecasts away from the professional consensus rather than toward it (Study 1). Generic chat did not add persuasion, but the chatbot's attitude mattered: participants whose chatbot pushed back moved 31% of the way toward its estimate, compared with 23% of the way when the chatbot sought common ground (Study 2). Combining the most persuasive feature levels into an optimized prompt almost doubled the weight on the chatbot's estimate relative to the generic baseline (39% vs. 23%) and left participants less uncertain about their revised beliefs (Study 3).

The conversation style mattered for how persuasive the bot was. In Study 1, generic chat did not increase persuasion beyond showing the AI's forecast and reasoning. In Study 3, by contrast, the optimized chatbot was substantially more persuasive than the generic baseline. Because every chatbot condition displayed an estimate and reasoning before a five-minute exchange, the gain came from the content and style of the conversation, not from interactivity alone. Human conversations cannot be engineered or standardized in this way. Chatbot conversations can, and our results show that this design choice matters.

The attitude result has two distinct design implications. First, commercial assistants are trained through human feedback to be helpful and agreeable, not specifically to correct beliefs or persuade users. That objective can reward sycophancy—the tendency to accommodate a user's stated view even when it is mistaken (Ouyang et al., 2022; Sharma et al., 2024). Study 2 shows the resulting tradeoff: a chatbot instructed to identify weak points, present counterarguments, and express disagreement was more persuasive than one instructed to find common ground and validate the participant's reasoning while maintaining its own forecast. The qualities that make an assistant pleasant and easy to talk to can therefore make it less effective at correcting beliefs.

Second, a chatbot may be an unusually effective way to deliver disagreement. People avoid disconfirming information and often respond negatively to opposing views from other people (Hart et al., 2009; Golman, Hagmann and Loewenstein, 2017; Yeomans et al., 2020; Minson and Chen, 2022). By contrast, counterattitudinal messages attributed to AI are perceived as less biased, more informative, and less intent on persuasion than the

same messages attributed to people, increasing receptiveness to them (Lu, Tormala and Duhachek, 2025). A challenging chatbot may therefore present arguments that users would resist or avoid in a human conversation.

Our accuracy results temper any optimism that AI influence naturally improves beliefs. In Study 1, the AI moved forecasts without moving them closer to professional expectations because its stated estimate was, by design, untethered. Nothing about a language model's persuasiveness guarantees the direction in which it persuades. A chatbot anchored to something else—engagement, an operator's interests, or deliberately seeded misinformation (Kreps, McCain and Brundage, 2022; Goldstein et al., 2024)—would move households with the same efficiency.

The uncertainty results add a second, less obvious dimension. The generic baseline prompt did not significantly reduce uncertainty relative to the control, whereas the optimized prompt did. For the economics of expectations, a point prediction alone does not reveal how firmly a belief is held (Manski, 2004; Engelberg, Manski and Williams, 2009). A persuasive AI can therefore both relocate and consolidate expectations. That is an opportunity for anchoring and a mechanism for entrenchment, depending entirely on what the system argues for.

Several limitations qualify our conclusions and chart a path for future work. First, our conversations were short, one-shot, and conducted with a single model, GPT-4o; whether repeated interactions amplify or erode the AI's advantage—and whether the challenge effect survives familiarity—remains open. Second, our participants were U.S. adults on Prolific, a comparatively attentive online pool (Peer et al., 2022), and their forecasts

were unincentivized, as household expectations surveys typically are. Third, our preregistered forecast-error outcome is not yet resolved: the forecast horizon extends to late 2026, so we cannot yet say whether AI persuasion left households closer to or farther from realized inflation. Finally, Study 3 validates the optimized *package* rather than each ingredient: several factor levels were selected based on point estimates that did not reach significance in Study 2, and the dimensions may interact in a larger study.

Managing inflation expectations has long relied on speeches, projections, and targets that reach households slowly and imperfectly. Conversations with trained, well-informed people offer a more direct channel, but such outreach is difficult to scale. Our results show that conversational AI can substantially shift beliefs, particularly when prompted to use an effective conversational style. Anchored to the professional consensus, conversational AI can therefore give central banks a scalable way to align household expectations with informed forecasts and strengthen the stabilizing effects of monetary policy.

REFERENCES

Anderson, Eric T., and Duncan I. Simester. 2010. "Price Stickiness and Customer Antagonism." *The Quarterly Journal of Economics*, 125(2): 729–765.

Andre, Peter, Ingar Haaland, Christopher Roth, Mirko Wiederholt, and Johannes Wohlfart. 2026. "Narratives about the Macroeconomy." *The Review of Economic Studies*, rdag014.

Armantier, Olivier, Wändi Bruine de Bruin, Giorgio Topa, Wilbert van der Klaauw, and Basit Zafar. 2015. "Inflation Expectations and Behavior: Do Survey Respondents Act on Their Beliefs?" *International Economic Review*, 56(2): 505–536.

Barrera, Oscar, Sergei Guriev, Emeric Henry, and Ekaterina Zhuravskaya. 2020. "Facts, Alternative Facts, and Fact Checking in Times of Post-Truth Politics." *Journal of Public Economics*, 182: 104123.

Belot, Michèle, and Guglielmo Briscese. 2026. "Breaking Echo Chambers: An Experimental Study on the Willingness to Listen to Opposite Views." *The Economic Journal*, ueag072.

Bick, Alexander, Adam Blandin, and David J. Deming. 2024. "The Rapid Adoption of Generative AI." National Bureau of Economic Research NBER Working Paper 32966, Cambridge, MA.

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): 425–457.

Bonaccio, Silvia, and Reeshad S. Dalal. 2006. "Advice Taking and Decision-Making: An Integrative Literature Review, and Implications for the Organizational Sciences." *Organizational Behavior and Human Decision Processes*, 101(2): 127–151.

Candia, Bernardo, Michael Weber, Yuriy Gorodnichenko, and Olivier Coibion. 2023. "Perceived and Expected Rates of Inflation of US Firms." *AEA Papers and Proceedings*, 113: 52–55.

Castelo, Noah, Maarten W. Bos, and Donald R. Lehmann. 2019. "Task-Dependent Algorithm Aversion." *Journal of Marketing Research*, 56(5): 809–825.

Coibion, Olivier, Dimitris Georgarakos, Yuriy Gorodnichenko, and Michael Weber. 2023. "Forward Guidance and Household Expectations." *Journal of the European Economic Association*, 21(5): 2131–2171.

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." *The Quarterly Journal of Economics*, 135(1): 165–219.

Coibion, Olivier, Yuriy Gorodnichenko, Saten Kumar, and Mathieu Pedemonte. 2020. "Inflation Expectations as a Policy Tool?" *Journal of International Economics*, 124: 103297.

Costello, Thomas H., Gordon Pennycook, and David G. Rand. 2024. "Durably Reducing Conspiracy Beliefs through Dialogues with AI." *Science*, 385(6714): eadq1814.

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

D'Acunto, Francesco, Ulrike Malmendier, Juan Ospina, and Michael Weber. 2021. "Exposure to Grocery Prices and Inflation Expectations." *Journal of Political Economy*, 129(5): 1615–1639.

Dietvorst, Berkeley J., Joseph P. Simmons, and Cade Massey. 2015. "Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err." *Journal of Experimental Psychology: General*, 144(1): 114–126.

Dräger, Lena, Michael J. Lamla, and Damjan Pfajfar. 2016. "Are Survey Expectations Theory-Consistent? The Role of Central Bank Communication and News." *European Economic Review*, 85: 84–111.

Duca-Radu, Ioana, Geoff Kenny, and Andreas Reuter. 2021. "Inflation Expectations, Consumption and the Lower Bound: Micro Evidence from a Large Multi-Country Survey." *Journal of Monetary Economics*, 118: 120–134.

Engelberg, Joseph, Charles F. Manski, and Jared Williams. 2009. "Comparing the Point Predictions and Subjective Probability Distributions

of Professional Forecasters." *Journal of Business & Economic Statistics*, 27(1): 30–41.

Feinberg, Matthew, and Robb Willer. 2015. "From Gulf to Bridge: When Do Moral Arguments Facilitate Political Influence?" *Personality and Social Psychology Bulletin*, 41(12): 1665–1681.

Freling, Traci H., Zhiyong Yang, Ritesh Saini, Omar S. Itani, and Ryan Rashad Abualsamh. 2020. "When Poignant Stories Outweigh Cold Hard Facts: A Meta-Analysis of the Anecdotal Bias." *Organizational Behavior and Human Decision Processes*, 160: 51–67.

Goldstein, Josh A, Jason Chao, Shelby Grossman, Alex Stamos, and Michael Tomz. 2024. "How Persuasive Is AI-generated Propaganda?" *PNAS Nexus*, 3(2): pgae034.

Golman, Russell, David Hagmann, and George Loewenstein. 2017. "Information Avoidance." *Journal of Economic Literature*, 55(1): 96–135.

Graeber, Thomas, Christopher Roth, and Florian Zimmermann. 2024. "Stories, Statistics, and Memory." *The Quarterly Journal of Economics*, 139(4): 2181–2225.

Granziera, Eleonora, Vegard H. Larsen, Greta Meggiorini, and Leonardo Melosi. 2025. "Speaking of Inflation: The Influence of Fed Speeches on Expectations." Norges Bank Working Paper 10/2025, Oslo.

Hackenburg, Kobi, Caroline Wagner, Luke Hewitt, Ben M. Tappin, Ed Saunders, Hannah Rose Kirk, Helen Margetts, and Christopher Summerfield. 2026. "AI Systems Out-Persuade Expert Humans."

Hagmann, David, Julia A. Minson, and Catherine H. Tinsley. 2024. "Personal Narratives Build Trust across Ideological Divides." *Journal of Applied Psychology*, 109(11): 1693–1715.

Hart, William, Dolores Albarracín, Alice H. Eagly, Inge Brechan, Matthew J. Lindberg, and Lisa Merrill. 2009. "Feeling Validated versus Being Correct: A Meta-Analysis of Selective Exposure to Information." *Psychological Bulletin*, 135(4): 555–588.

Harvey, Nigel, and Ilan Fischer. 1997. "Taking Advice: Accepting Help, Improving Judgment, and Sharing Responsibility." *Organizational Behavior and Human Decision Processes*, 70(2): 117–133.

Huang, Karen, Michael Yeomans, Alison Wood Brooks, Julia Minson, and Francesca Gino. 2017. "It Doesn't Hurt to Ask: Question-asking Increases Liking." *Journal of Personality and Social Psychology*, 113(3): 430–452.

Hussein, Mohamed A., and Zakary L. Tormala. 2021. "Undermining Your Case to Enhance Your Impact: A Framework for Understanding the Effects of Acts of Receptiveness in Persuasion." *Personality and Social Psychology Review*, 25(3): 229–250.

Kalla, Joshua L., and David E. Broockman. 2020. "Reducing Exclusionary Attitudes through Interpersonal Conversation: Evidence from Three Field Experiments." *American Political Science Review*, 114(2): 410–425.

Kalla, Joshua L., and David E. Broockman. 2023. "Which Narrative Strategies Durably Reduce Prejudice? Evidence from Field and Survey Experiments Supporting the Efficacy of Perspective-Getting." *American Journal of Political Science*, 67(1): 185–204.

Karmarkar, Uma R., and Zakary L. Tormala. 2010. "Believe Me, I Have No Idea What I'm Talking About: The Effects of Source Certainty on Consumer Involvement and Persuasion." *Journal of Consumer Research*, 36(6): 1033–1049.

Kreps, Sarah, R. Miles McCain, and Miles Brundage. 2022. "All the News That's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation." *Journal of Experimental Political Science*, 9(1): 104–117.

Logg, Jennifer M., Julia A. Minson, and Don A. Moore. 2019. "Algorithm Appreciation: People Prefer Algorithmic to Human Judgment." *Organizational Behavior and Human Decision Processes*, 151: 90–103.

Løhre, Erik, Subramanya Prasad Chandrashekar, Lewend Mayiwar, and Thorvald Hærem. 2024. "Uncertainty, Expertise, and Persuasion: A Replication and Extension of Karmarkar and Tormala (2010)." *Journal of Experimental Social Psychology*, 113: 104619.

Lu, Louise, Zakary L. Tormala, and Adam Duhachek. 2025. "How AI Sources Can Increase Openness to Opposing Views." *Scientific Reports*, 15(1): 17170.

Malmendier, Ulrike, and Stefan Nagel. 2016. "Learning from Inflation Experiences." *The Quarterly Journal of Economics*, 131(1): 53–87.

Manski, Charles F. 2004. "Measuring Expectations." *Econometrica*, 72(5): 1329–1376.

Minson, Julia A., and Frances S. Chen. 2022. "Receptiveness to Opposing Views: Conceptualization and Integrative Review." *Personality and Social Psychology Review*, 26(2): 93–111.

Oba, Demi, Jonah Berger, and Reihane Boghrati. 2026. "Effectively Communicating Uncertainty: The Persuasive Impact of Different Types of Hedges." *Journal of Consumer Psychology*, 36(3): 314–332.

Ouyang, Long, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. "Training Language Models to Follow Instructions with Human Feedback." 27730–27744. New Orleans, Louisiana, USA:Neural Information Processing Systems Foundation, Inc. (NeurIPS).

Peer, Eyal, David Rothschild, Andrew Gordon, Zak Evernden, and Ekaterina Damer. 2022. "Data Quality of Platforms and Panels for Online Behavioral Research." *Behavior Research Methods*, 54(4): 1643–1662.

Petty, Richard E., and John T. Cacioppo. 1984. "The Effects of Involvement on Responses to Argument Quantity and Quality: Central and Peripheral Routes to Persuasion." *Journal of Personality and Social Psychology*, 46(1): 69–81.

Pilossoph, Laura, and Jane M. Ryngaert. 2024. "Job Search, Wages, and Inflation." National Bureau of Economic Research NBER Working Paper 33042, Cambridge, MA.

Price, Paul C., and Eric R. Stone. 2004. "Intuitive evaluation of likelihood judgment producers: evidence for a confidence heuristic." *Journal of Behavioral Decision Making*, 17(1): 39–57.

Sah, Sunita, Don A. Moore, and Robert J. MacCoun. 2013. "Cheap Talk and Credibility: The Consequences of Confidence and Accuracy on Advisor Credibility and Persuasiveness." *Organizational Behavior and Human Decision Processes*, 121(2): 246–255.

Salvi, Francesco, Manoel Horta Ribeiro, Riccardo Gallotti, and Robert West. 2025. "On the Conversational Persuasiveness of GPT-4." *Nature Human Behaviour*, 9(8): 1645–1653.

Sharma, Mrinank, Meg Tong, Tomasz Korbak, David Duvenaud, Amanda Askell, Samuel R. Bowman, Newton Cheng, Esin Durmus, Zac Hatfield-Dodds, Scott R. Johnston, Shauna M. Kravec, Timothy Maxwell, Sam McCandlish, Kamal Ndousse, Oliver Rausch, Nicholas Schiefer, Da Yan, Miranda Zhang, and Ethan Perez. 2024. "Towards Understanding Sycophancy in Language Models."

Soll, Jack B., and Richard P. Larrick. 2009. "Strategies for Revising Judgment: How (and How Well) People Use Others' Opinions." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3): 780–805.

Voelkel, Jan G., Michael N. Stagnaro, James Y. Chu, Sophia L. Pink, Joseph S. Mernyk, Chrystal Redekopp, Isaias Ghezae, Matthew Cashman, Dhaval Adjodah, Levi G. Allen, L. Victor Allis, Gina Baleria, Nathan Ballantyne, Jay J. Van Bavel, Hayley Blunden, Alia Braley, Christopher J. Bryan, Jared B. Celniker, Mina Cikara, Margarett V. Clapper, Katherine Clayton, Hanne Collins, Evan DeFilippis, Macrina Dieffenbach, Kimberly C. Doell, Charles Dorison, Mylien Duong, Peter Felsman, Maya Fiorella, David Francis, Michael Franz, Roman A. Gallardo, Sara Gifford, Daniela Goya-Tocchetto, Kurt Gray, Joe Green, Joshua Greene, Mertcan Güngör, Matthew Hall, Cameron A. Hecht, Ali Javeed, John T. Jost, Aaron C. Kay, Nick R. Kay, Brandyn Keating, John Michael Kelly, James R. G. Kirk, Malka Kopell, Nour Kteily, Emily Kubin, Jeffrey Lees, Gabriel Lenz, Matthew Levendusky, Rebecca Littman, Kara Luo, Aaron Lyles, Ben Lyons, Wayde Marsh, James Martherus, Lauren Alpert Maurer, Caroline Mehl, Julia Minson, Molly Moore, Samantha L. Moore-Berg, Michael H. Pasek, Alex Pentland, Curtis Puryear, Hossein Rahnama, Steve Rathje, Jay Rosato, Maytal Saar-Tsechansky, Luiza Almeida Santos, Colleen M. Seifert, Azim Shariff, Otto Simonsson, Shiri Spitz Siddiqi, Daniel F. Stone, Palma Strand, Michael Tomz, David S. Yeager, Erez Yoeli, Jamil Zaki, James N. Druckman, David G. Rand, and Robb Willer. 2024. "Megastudy Testing 25 Treatments to Reduce Antidemocratic Attitudes and Partisan Animosity." *Science*, 386(6719): eadh4764.

Weaver, Kimberlee, Stefan J. Hock, and Stephen M. Garcia. 2016. ""Top 10" Reasons: When Adding Persuasive Arguments Reduces Persuasion." *Marketing Letters*, 27(1): 27–38.

Weber, Michael, Francesco D'Acunto, Yuriy Gorodnichenko, and Olivier Coibion. 2022. "The Subjective Inflation Expectations of Households and Firms: Measurement, Determinants, and Implications." *Journal of Economic Perspectives*, 36(3): 157–184.

Wood, Wendy, Carl A Kallgren, and Rebecca Mueller Preisler. 1985. "Access to Attitude-Relevant Information in Memory as a Determinant of Persuasion: The Role of Message Attributes." *Journal of Experimental Social Psychology*, 21(1): 73–85.

Yeomans, Michael, Julia Minson, Hanne Collins, Frances Chen, and Francesca Gino. 2020. "Conversational Receptiveness: Improving Engagement with Opposing Views." *Organizational Behavior and Human Decision Processes*, 160: 131–148.

Zebregs, Simon, Bas van den Putte, Peter Neijens, and Anneke de Graaf. 2015. "The Differential Impact of Statistical and Narrative Evidence on Beliefs, Attitude, and Intention: A Meta-Analysis." *Health Communication*, 30(3): 282–289.

Appendix

Chatbot Prompts

Baseline prompt (Studies 1–3). The AI partner received the participant's and its own forecast, reasoning, and information sources, followed by these guidelines (used in Study 1 and, with the factor modifications below, in Studies 2–3):

> Be conversational and friendly, like a real participant. Actively share your perspective, observations, and reasoning rather than only asking questions. Reference specific examples from both your sources and the participant's, and draw connections between them. Balance sharing your views (60%) with asking about theirs (40%). Explain *why* you hold your forecast, citing specific sources, and engage with the participant's sources. Keep responses concise but substantive (2–4 sentences). Be respectful, avoid repetition, draw on a broad set of factors, and reference everyday observations such as shopping prices, rent, wages, and news. Do not use technical economic jargon; write like an everyday person, not an economist.

Study 2 factor modifications. Each factor appended one instruction, which overrode the corresponding baseline guidance (Table 2).

Study 3 optimized prompt. The optimized prompt is the baseline prompt with the most persuasive level of each factor appended. The added instructions, relative to the baseline above, are listed below:

- Write detailed responses of 5–7 sentences with multiple supporting

points. (*length: long*)

- Express your views with high confidence ("I'm certain", "definitely", "there's no doubt"); avoid hedging. (*certainty: high*)
- Support your arguments with concrete everyday examples people experience—grocery prices, rent, gas—described as general observations; do not cite statistics. (*information: examples*)
- Always include at least one question in each response to engage the participant. (*questions: yes*)
- Firmly and directly challenge the participant's perspective: identify weak points and inconsistencies, present counterarguments, and express disagreement ("I disagree", "I don't think this is right"); use no affirming or validating language. (*attitude: challenging*)

Study 3 control prompt. The control chatbot was instructed to discuss information sources and to avoid the forecast itself:

> You are a participant in a research study having a friendly discussion about where people get their information from and why different people may look at different information sources.
>
> TOPIC: Information sources and media consumption
>
> - Discuss where people get their news and information (social media, TV, newspapers, podcasts, friends, etc.)
> - Explore why different people might prefer different sources
> - Talk about how accessing different information sources can lead to different views on topics

- Be curious about where the participant gets their information and why

IMPORTANT RULES:

- Do NOT discuss inflation, economics, or prices specifically
- If the participant mentions inflation or economics, acknowledge briefly but redirect the conversation to the broader question of information sources and how they shape views generally
- Focus on the meta-level question of WHY people choose different sources, discuss the information-source choices on non-economic topics
- Be warm, conversational, and thoughtful
- When discussing information sources, talk about general patterns and ask about the participant's preferences
- Keep responses concise but substantive (2-4 sentences typically)

Start by asking where they typically get their news and information from.