

Inflation Expectations & Central Bank Communication



Alena Wabitsch

St Catherine's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Economics

Hilary 2025

Abstract

This thesis studies how households form and report inflation expectations, and how communication can be used as a policy tool to guide these macroeconomic expectations. I use a combination of big data, experiments, and theoretical modeling to address critical questions: Is central bank communication successfully reaching the public? What is the role of the messenger for this communication? And does it matter how macroeconomic expectations are elicited?

Chapter 1, *“The Messenger Matters”*, explores the role the messenger plays in reaching the public and shaping individuals’ beliefs about future inflation. To investigate how messenger characteristics affect central bank communication, focusing on nationality within the Euro area, I first create a novel dataset of over eight million tweets in five languages, and then design a tailored inflation forecasting experiment that I implement in four different countries. I find positive ingroup effects: communication is more effective—in both reach and influence—when the characteristics of the messenger and receiver align. The causal ingroup effect on inflation expectations is driven by homophily and rooted in trust. Delegating communication to national central banks or other board members replicates these positive effects. Based on these novel empirical insights, I develop a stylized coordination model that demonstrates how strategically selecting messengers maximizes social welfare, thereby offering a powerful policy tool for effective public communication. This work was awarded the *Peter Sinclair Prize* from the *British Money Macro and Finance Society (MMF)*.

Chapter 2, *“Central Bank Communication With Non-Experts—A Road to Nowhere?”*, explores whether central banks successfully reach the public with their efforts to communicate effectively to manage inflation expectations, foster trust, and reduce uncertainty. We find that non-experts are listening occasionally, but experts are reached much more easily. This chapter is co-authored with Michael Ehrmann and published in the *Journal of Monetary Economics*.

Finally, Chapter 3, *“Incentivizing Inflation Expectations”*, questions the common methodology used to elicit household inflation expectations, which are crucial for economic modeling and policymaking. Despite the well-established role of incentives in experimental economics, all major surveys of inflation expectations pay a

flat participation fee. This lack of marginal incentives extends to many information provision experiments—often designed as randomized controlled trials (RCTs). We introduce marginal incentives into a standard survey of inflation expectations, and find that marginal incentives significantly alter expectation distributions. Reported mean forecasts are lower, and cross-sectional disagreement diminishes. We even find that marginal incentives eliminate the gender expectations gap, a long-standing puzzle in the literature on expectations. Furthermore, in an embedded RCT, marginal incentives lead to greater responsiveness to information provision, contrasting with null effects under flat fees in times when attention to inflation is already heightened. These findings underscore the importance of marginal incentives in surveys and survey-based experiments to enhance data validity, strengthen empirical research, and better inform policymaking. This chapter is co-authored and benefited from competitive funding through a *British Academy/Leverhulme Small Research Grant*.

Supervisors

This thesis was supervised by Michael McMahon and Alexandre Kohlhas, both from the University of Oxford.

Total Words

This thesis contains approximately 70,000 words.

Declarations and acknowledgment of funding

No chapter in this thesis has been submitted for a degree at this or any other university. I acknowledge funding from the Austrian Economic Association (NOeG), the British Academy/Leverhulme, the Department of Economics (Oxford), St Catherine's College (Oxford), the George Webb Medley Fund, and the Oxford Economics Papers Fund. All experiments in this thesis have obtained prior ethical approval from the University of Oxford CUREC committee.

Alena Wabitsch (*March 2025*)

Acknowledgments

First, I would like to thank my advisors Michael McMahon and Alexandre Kohlhas; your support made all the difference.

Michael, thank you for your wisdom, your intuition and your guidance. Your iconic stories taught me a lot about economic research and encouraged me to be brave with my ideas.

Alex, I am lucky to have experienced your brilliance and your remarkable ability to make everything simple. Thank you for being such an endless source of inspiration and for continuously challenging me to be a better researcher.

I am also grateful to my co-authors, who taught me so much about economics, research, and life. Michael Ehrmann, thank you for teaching me how to conduct research and for your incredible patience every step of the way. Ryan Rholes, thank you for sharing your expertise and your excitement for experimental macroeconomics. Tobias Cagala, thank you for introducing me to coding, and for making research fun. Ana Lamo, Antonio Dias da Silva, JF Jimeno and Stefania Albanesi, thank you for your dedication and support. Tanja Linta and Manuel Mosquera-Tarrio, thank you for facing challenges with me and for growing together. You greatly shaped my DPhil journey.

I would like to thank Oxford's Macro Group for enriching years of discussions and insightful exchanges about economics, and for sharing the joys of research.

I would also like to thank Severine Toussaerts and the other members of Oxford's Behavioural Lab for their invaluable insights about experimental design and microeconomic theory.

I am also grateful to my parents, siblings and friends for their love and support, and for their voluntary and involuntary participation in testing all my experiments and theories.

Finally, I would like to thank Edvard for sharing the high points as well as the low points.

Contents

1	The Messenger Matters	1
1.1	Introduction	2
1.1.1	Related Literature	4
1.2	Three Stylized Facts	6
1.2.1	Twitter Data	6
1.2.2	Stylized Fact 1: Higher Reporting on Ingroup Policymakers	7
1.2.3	Stylized Fact 2: Higher Information Availability	8
1.2.4	Stylized Fact 3: Stronger Belief Updating	9
1.2.5	Discussion	11
1.3	Inflation Forecasting Experiment	12
1.3.1	Experimental Design	12
1.3.2	Experimental Sample	16
1.3.3	Estimating Belief Updating	17
1.3.4	Causal Findings on Updating Inflation Expectations	18
1.3.5	Mechanism: Perceived Quality and Trust	22
1.3.6	Ingroup Policymakers Raise Information Reach but Not At- tention	23
1.3.6.1	Attention Is Endogenous	23
1.3.6.2	Attention Raises Ingroup Effects	23
1.3.7	Discussion of Experimental Findings and Policy Implications	24
1.4	Modeling Optimal Communication	25
1.4.1	The Model Set-Up	26
1.4.1.1	Information Structure	26
1.4.1.2	Payoffs	28
1.4.1.3	Social Welfare	28
1.4.1.4	Agents' Actions	28

1.4.1.5	Timeline	29
1.4.2	Equilibrium	29
1.4.2.1	Derivation of Equilibrium Actions	29
1.4.2.2	Uniqueness of Equilibrium	30
1.4.3	Welfare Implications and Policy Analysis	30
1.4.3.1	Social Welfare Computation	31
1.4.3.2	The Precision of Public Information: The Disclosure Policy	31
1.4.3.3	The Composition of In- and Outgroup Agents: The Delegation Policy	33
1.4.4	Policy Implications: When to Delegate?	35
1.5	Conclusion	37
2	Central Bank Communication With Non-Experts—A Road to Nowhere?	39
2.1	Introduction	39
2.2	Related Literature	41
2.3	Data	43
2.3.1	Tweets	43
2.3.2	Content of Tweets	44
2.3.3	Associated Twitter Accounts	46
2.3.4	ECB Communication Events	46
2.4	Differentiating Experts from Non-Experts	47
2.5	Determinants of Retweets, Likes and Replies	50
2.6	Responsiveness of Tweets to ECB Communication	52
2.6.1	Hypotheses and Specification of the Econometric Model	52
2.6.2	Twitter Traffic and User Concentration	54
2.6.3	The Content of Tweets	55
2.6.4	Any Difference for Tweets in German?	57
2.7	Conclusion	59
3	Incentivizing Inflation Expectations	61
3.1	Introduction	61
3.2	Experimental Design	64
3.2.1	Hypotheses	66
3.2.2	Data	67
3.3	Results	67
3.3.1	The Effect of Incentives on Elicited Expectations	67

3.3.2	The Effect of Incentives on (In)Attention and Backward-Looking Heuristics	71
3.3.3	Incentives and the Puzzle of the Gender Gap in Inflation Expectations	73
3.3.4	Effect of Incentives in Information Provision Experiments . . .	75
3.3.5	Incentives and Effort	77
3.4	Discussion	79
A	Appendix to Chapter 1: The Messenger Matters	82
A.1	Details on Twitter Data and Additional Evidence	82
A.1.1	Cleaning Twitter Data	82
A.1.2	Descriptive Analysis of Twitter Data	82
A.1.3	Additional Tables and Figures	86
A.1.4	Does Print Media Confirm Increased Information Availability? . . .	87
A.1.5	Empirical Estimation of Belief Updating	90
A.2	Additional Information on the Experiment	95
A.2.1	Power Analysis	99
A.2.2	Bayesian Belief Updating Framework	99
A.2.3	Additional Experimental Results, Tables and Figures	100
A.2.4	Information Availability: Exposure to News in the Media . . .	105
A.2.5	Real-World Forecasting	106
A.3	Interface of the Experiment	107
A.4	Model Details	121
A.4.1	Threshold for Public Signal Availability	121
A.4.2	Proofs and Derivations	121
A.4.2.1	Proof of Lemma 1: Optimal Signal Extraction Weights	121
A.4.2.2	Proof of Proposition 1: Uniqueness of Linear Equilibrium	123
A.4.2.3	Private Signal Precision May Harm Welfare	124
A.4.2.4	Derivation for Proposition 2: Welfare Implications of the Public Signal Precision	125
A.4.2.5	Derivation for Proposition 3: The Optimal Share of Ingroup Agents	126
A.4.3	Details on Expertise Loss	126
A.4.3.1	Proof Sketch of Proposition 5: Delegating Communication vs. Loss in Expertise	126

B	Appendix to Chapter 2: Central Bank Communication With Non-Experts— A Road to Nowhere?	129
B.1	Details on the Data Construction	129
B.1.1	Cleaning Procedure for ECB-Related Tweets	129
B.1.2	Summary Statistics	131
B.2	Robustness of Results	135
C	Appendix to Chapter 3: Incentivizing Inflation Expectations	144
C.1	Other Tables and Figures	144
C.2	Power Analysis	148
C.3	Inflation Expectations Survey	149

List of Figures

1.1	Reporting on Policymakers by Language	8
1.2	Share of Tweets by Language	9
1.3	Belief Updates by Tweet Language	11
1.4	Timeline of Experiment	12
1.5	Signal Use Across Treatments by Nationality	18
1.6	Disclosure's Effect on Social Welfare	32
1.7	Optimal Share of Ingroup Agents	34
1.8	Social Welfare Contours	35
2.1	Interest in the ECB: Number of Tweets, Google Searches, Newspaper Articles	45
2.2	Lorenz Curve of Twitter Activity by Events	55
3.1	CDFs of Expected Inflation By Treatment	68
3.2	Hypothetical Payoffs for Point Forecasts of Inflation	71
3.3	Average Bin Weights across Incentives	78
A.1	Daily Tweet Volume by Language	83
A.2	Word Clouds by Language	86
A.3	National Newspaper Articles Over Time	88
A.4	Distributions of Priors and Posteriors	90
A.5	Belief Updating after Press Conferences By Expertise	93
A.6	Quarterly EA Inflation, SPF Forecasts, and SPF Forecast Errors	95
A.7	Inflation Sequences	96
A.8	Power Analysis	100
A.9	Descriptive Evidence of Experiment Participants	101
A.10	Requesting Additional Information Pieces (With 95%-CI)	103
A.11	Knowing Representative Policymakers and Receiving News	105

A.12 Comparing Household Inflation Forecasts to ECB Projections	106
A.13 Welcome Page	107
A.14 Instructions (1)	108
A.15 Instructions (2)	109
A.16 Comprehension Quiz	110
A.17 Prior Elicitation	111
A.18 Posterior Elicitation	112
A.19 Attention Check	113
A.20 Survey: Purpose of Experiment & Strategy	113
A.21 Survey: Association	114
A.22 Survey: Ability	115
A.23 Survey: Exposure	116
A.24 Survey: Trust	117
A.25 Survey: Monetary Policy Expertise	118
A.26 Payoff Overview	119
A.27 Payoff Reveal	120
A.28 Social Welfare Impact of Delegation with Expertise Loss	128
 B.1 Word Clouds of ECB-Related and ECB-Unrelated Tweets	 130
 C.1 CDFs of Inflation Expectations After Information Provision	 145
C.2 Welcome	149
C.3 General Questions	150
C.4 Explanations	150
C.5 Inflation Point Forecast (Flat)	151
C.6 Inflation Point Forecast (Posterior)	152
C.7 Inflation Point Forecast (Prior)	153
C.8 Inflation Point Forecast (Both)	154
C.9 Food Point Forecast	155
C.10 Gas Point Forecast	155
C.11 Information Intervention	156
C.12 Food Bin Forecast	156
C.13 Gas Bin Forecast	157
C.14 Inflation Bin Forecast (Flat and Prior)	158
C.15 Inflation Bin Forecast (Posterior)	159
C.16 Inflation Bin Forecast (Both)	160
C.17 End of Survey	161

List of Tables

1.1	Overview of Messenger Treatments	14
1.2	Main Results from Experiment (H1, H2, H3, H4)	20
1.3	Endogenous Attention	24
2.1	Summary Statistics For Different Account Types	49
2.2	Determinants of Retweets, Likes and Replies	51
2.3	Explaining Twitter Traffic and Concentration of Users	56
2.4	Explaining Subjectivity	57
2.5	Explaining Favourableness	58
2.6	Explaining Absolute Favourableness	59
3.1	Overview of Treatments	66
3.2	Summary Statistics and Variance Comparison of Inflation Expectations	69
3.3	Effects of Incentives on Inflation Expectations	70
3.4	Comparing Experimental Data to SCE and SPF	71
3.5	Effects of Incentives on Inattention and Perception	72
3.6	Effects of Incentives on the Gender Expectations Gap	74
3.7	Effect of Incentives on Learning Rates	76
3.8	Effect of Incentives on Effort	79
A.1	Peaks in Tweet Volume by Language	84
A.2	Sentiment of Tweets by Language	85
A.3	Number of Tweets and Distinct Twitter Users	85
A.4	Ingroup Effect on Information Supply on Twitter	87
A.5	Effect of ECB President's Nationality on Newspaper Volume	89
A.6	Belief Updates at Different Points in Time	93
A.7	Belief Updating after Press Conferences	94
A.8	Overview of Randomly Selected Sequences	95

A.9	Additional Information Pieces and Corresponding Inflation Scenarios	98
A.10	Representative Policymakers by Nationality	99
A.11	Regression Results By Nationality	101
A.12	Mechanism: Perceived Quality and Trust	102
A.13	The Role of Attention and Being Informed	104
A.14	Information Availability and Reach: Survey Evidence	105
B.1	Number of ECB-Related Tweets and Retweets	131
B.2	Descriptive Statistics of Tweet Content	132
B.3	Examples of Words in English Sentiment Lexicon	132
B.4	Characteristics of Account Types, Robustness	133
B.5	Twitter Traffic, Leads and Lags of Press Conference and “Whatever It Takes”	134
B.6	Twitter Traffic, Robustness Tests	136
B.7	Concentration Index, Robustness Tests	137
B.8	Average Subjectivity, Robustness Tests	138
B.9	Standard Deviation of Subjectivity, Robustness Tests	139
B.10	Average Favourableness, Robustness Tests	140
B.11	Standard Deviation of Favourableness, Robustness Tests	141
B.12	Average Absolute Favourableness, Robustness Tests	142
B.13	Standard Deviation of Absolute Favourableness, Robustness Tests . .	143
C.1	Sample Comparisons: Across Groups and SCE	144
C.2	Hypothetical Earnings From Inflation Expectations (The Priors) . . .	145
C.3	Summary Statistics and Variance Comparison of APE	146
C.4	Effects of Incentives on Updated Inflation Expectations (The Posteriors)	146
C.5	Effect of Incentives on Effort	147
C.6	Sample Size Calculation	148

The Messenger Matters*

Abstract Does the messenger matter in communication with the public? For policy communication on monetary, climate, fiscal, or other issues to be impactful, it must successfully reach people and influence their beliefs. I combine empirical and theoretical evidence to explore how messenger characteristics affect central bank communication, focusing on nationality in the Euro area. Leveraging big data from Twitter and a tailored inflation forecasting experiment, I find positive ingroup effects: communication is more effective—in both reach and influence—when the messenger and receiver share nationality. Delegating communication to national central banks or other board members can replicate these benefits. Using a stylized coordination model, I identify that strategically selecting messengers maximizes social welfare, offering a novel, powerful policy tool for effective public communication, complementing traditional disclosure policies.

Keywords: Central Bank Communication, Inflation Expectations, Ingroup Effects, Coordination Games

JEL: E52, E58, D82, D83, D84

*This chapter is based on Wabitsch (2024). Ethical approval: ECONCIA21-22-24. AEA RCT Registry ID: AEARCTR-0010727. Funded by: Austrian Economic Association (NOeG), St Catherine's College (Oxford), and Department of Economics (Oxford). For valuable comments I am grateful to Miguel Ballester, Tobias Cagala, Michael Ehrmann, Basil Halperin, Christian Hellwig, Alexandre Kohlhas, Seung Lee, Alistair Macaulay, Michael McMahon, Matthew Naylor, Lukas Nord, Luba Petersen, Ryan Rholes, Peter Rickards, Hannah Roemer, Séverine Toussaert, as well as for participants at the following seminars and conferences: 2025: Cleveland Fed, WU Vienna, University of Amsterdam, University of Mannheim, Universitat Autònoma de Barcelona, University of Manchester, Nova SBE, Norwegian School of Economics (NHH); 2024: CEBRA Annual Meeting, EEA-ESEM, EABCN: New Challenges in Monetary Economics (University of Mannheim), Behavioral and Experimental Economics Workshop (Esade), EAYE (Paris School of Economics), Annual MMF PhD Conference (University of Surrey), CEPR Central Bank Communication RPN, London School of Economics (Behavioural Finance Group), Oxford-CEPR Central Bank Communication Workshop, University of Oxford (Macroeconomics Workshop); 2023: Austrian Economic Association (NOeG) Winter Workshop, University of Oxford (Behavioural Lab), the Annual Meeting of the NOeG, Sveriges Riksbank, Expectations in Dynamic Macroeconomic Models (TU Vienna).

1.1 Introduction

Effective communication of a policy to the public is critical for its impact: people will only respond if they receive the information and believe it. This includes health campaigns promoting vaccine uptake during pandemics, subsidies and taxes incentivizing climate-friendly behavior, and monetary policy aimed at influencing consumption and savings decisions. How can policies—whether on climate, health, fiscal, monetary, and other issues—be communicated effectively to the public?

One important dimension that can be varied is the messenger. Communicating with the general public means that the audience is heterogeneous across various dimensions—nationality, gender, education, and more. Literature on social identity has identified “in-group effects”: Receivers respond more strongly to messages delivered by someone who shares characteristics with them (i.e., someone from an “ingroup”), as formalized for economic decision-making by Malmendier and Veldkamp (2022). Since policy communication occurs in institutional context, the identity of messengers has widely been neglected by policymakers and the policy communication literature. However, should ingroup effects also exist for policy communication, choosing diverse messengers could make this communication more effective.

In this paper, I combine novel empirical evidence and theoretical methods to address two questions. First, positively, how do messenger characteristics impact policy communication with the public? Second, normatively, how should policy be communicated to the public in light of these messenger effects? I explore this in the context of central bank communications, as central banks have made communication a core component of their monetary policy toolkit, recognizing its power (see Blinder et al. (2024) for a recent overview). Specifically, to investigate ingroup dynamics, I study how nationality—a salient yet understudied characteristic—influences central bank communication in the Euro area. First, I create a new multilingual dataset of tweets that provides novel evidence of ingroup effects along two dimensions: information availability and information processing (belief updating). Second, I design a new inflation forecasting experiment, in which participants of four different nationalities forecast inflation subject to receiving signals from various messengers. The treatment is an isolated variation in messenger nationality, allowing me to precisely estimate the pure causal effect of messenger nationality on inflation expectations. With this design, I further confirm the efficacy of two concrete examples of delegating communication: to other board members of the same institution, or to other institutions. Finally, accounting for these reduced-form insights, I develop a stylized coordination model to derive optimal, welfare-maximizing communication policies. I find that strategic delegation of communication is a new and powerful policy tool that complements the well-studied disclosure policy. When the quality of public information is high, delegating to heterogeneous messengers maximizes welfare; conversely, centralizing communication can be more socially desirable when public information quality is low.

Three novel stylized facts based on big data from Twitter motivate that the messenger’s nationality matters in the monetary union of the Euro. First, the reporting on the messenger (here, the president of the European Central Bank (ECB)) rises strongly for the ingroup.

Second, information availability—measured as the volume of tweets—is higher for the ingroup. And third, after the ECB’s press conference, Twitter users update their beliefs more strongly when they are in the ingroup. These facts are based on over 8 million multilingual tweets posted between 2016 and 2022, which give meaningful insight into real-world phenomena such as modern information supply and responses to communication events.

To establish causal mechanisms, I design a pre-registered inflation forecasting experiment tailored to assess ingroup effects on inflation expectations, underlying mechanisms, and messenger-driven attention to information. In this experiment, participants from Germany, Spain, France, and Italy forecast inflation under six different scenarios. Each time, they receive a signal about future inflation from varying messengers, and then get the chance to update their forecasts. The treatment consists of isolated variation in the nationality of otherwise generic messengers. Using a standard Bayesian belief updating framework, I estimate the causal nationality-based ingroup effect on updating inflation expectations. The within-subject design allows me to further distinguish between two drivers of ingroup effects: homophily (preference for similar others) and heterophobia (aversion to different others). In addition, I provide causal reduced-form evidence for two concrete policy recommendations: delegating communication to diverse messengers within an institution, and delegating communication to other institutions.

Across all messenger treatments, participants under-use signals by approximately ten percentage points compared to a Bayesian benchmark. Signals from ingroup messengers are used 5.2 percentage points more than those from outgroup messengers—roughly half of the average information under-use. This constitutes the pure nationality-based ingroup effect on information processing related to inflation expectations. Delegating communication to ECB experts of the same nationality, or to national central banks (NCBs), successfully replicates these ingroup effects, albeit to a lesser extent. The positive effects on belief updating are primarily driven by trust, largely associated with reported perceived quality, and homophily rather than heterophobia.

After the experiment, participants face a survey, which provides more precise insights on information availability, complementing Twitter findings. Individuals indeed receive more news about their respective ingroup policymakers, confirming better reach beyond information availability. However, while messengers influence real-world information availability, they do not causally affect attention to information in the experiment. Instead, attention responds causally to the level of inflation and its uncertainty, in line with attention to information being endogenous to the inflationary environment.

The evidence from both Twitter and the experiment shows that messenger effects positively impact the extent to which the ingroup receives and uses information. This makes central bank communication more effective for ingroup audiences. Reduced-form evidence further underscores benefits of delegating communication to board members of certain nationality or to the Eurosystem of NCBs. In light of these messenger effects, what should optimal public communication policy look like?

To formally assess the normative predictions for optimal communication policy that maximizes social welfare, I develop a model in the spirit of a “beauty contest” (Morris and Shin, 2002). Agents are either in the in- or outgroup with the messenger, and they care about aligning their actions with an unknown fundamental as well as with all other agents’ actions. While linked to empirical findings based on nationality, the model remains general,

allowing findings to extend to public communication with heterogeneous messengers and receivers across various dimensions (such as gender, ethnicity, socio-economic background, or economic expertise), applicable to policy communication beyond monetary issues. Two communication policies are evaluated with the model: a disclosure policy and a delegation policy. The disclosure policy involves setting the precision of public information, while the delegation policy entails choosing the messenger(s) to deliver public information, thereby influencing the share of agents classified as ingroup.

I find that when the coordination motive is weak, increasing the share of ingroup agents through delegation is always beneficial—particularly so when public information is relatively precise. However, when coordination is strong and disclosing public information harms welfare because agents over-rely on imprecise public information and neglect their better private information, the presence of outgroup agents can mitigate such welfare losses. This is because outgroup agents receive and use less public information. Additionally, as the quality of public information improves, disclosure improves welfare, while delegation may still be harmful because in- and outgroup agents use public signals differently. Thus, strategic delegation—beyond disclosure—serves as a powerful additional policy tool to maximize welfare.

The messenger matters. For effective policy communication, the characteristics of the messenger and receivers are important. By strategically delegating communication to messengers with diverse characteristics, policymakers can influence how messages are received and processed, and ultimately maximize social welfare through their disclosure of public information.

1.1.1 Related Literature

Central bank communication with the general public While communication with financial market participants has been an acknowledged powerful policy tool for a while now, targeting the non-expert public is a more recent endeavor of central banks. Although deemed a potentially powerful policy tool affecting institutional trust, inflation expectations, and economic volatility after monetary policy decisions, doubts remain whether central banks actually reach the public. Blinder et al. (2024) summarize this literature comprehensively, building on work by e.g., Blinder et al. (2008); Ehrmann et al. (2013); Binder (2017); Blinder (2018); Haldane and McMahon (2018); Coibion et al. (2020). Central bank communication successfully affects media coverage (Ter Ellen et al., 2022; Lamla and Vinogradov, 2019), and although non-experts do respond to some communication events, they are generally not as responsive as experts (Ehrmann and Wabitsch, 2022). I add to this literature by evaluating the importance of the messenger for reaching non-experts, and by providing novel positive and normative evidence on concrete recommendations for optimal communication policy.

Ingroup effects and social identity Social identity influences various domains, including economics, education, healthcare, and finance. In the context of central banking, recent evidence suggests policymakers' gender and ethnicity affect their perception, trust, and influence (D'Acunto et al., 2022; Bodea et al., 2021; Bodea and Kerner, 2022).¹ Mal-

¹In education, studies show that Black teachers positively impact Black students' educational outcomes, with stronger effects when gender also matches (Gershenson et al., 2022; Price, 2010;

mendier and Veldkamp (2022) formalize ingroup-dependent belief formation in economic decision-making more generally. I am the first to incorporate this into a model of strategic complementarities, after empirically estimating it in the context of central bank communication. My estimation of the nationality-related ingroup effect on belief updating represents novel causal evidence attributed to a single, isolated characteristic. With my unique experimental design, I can precisely identify underlying mechanisms, and provide novel evidence on concrete policy recommendations.

Experiments in Macroeconomics I design a new inflation forecasting experiment to causally infer the effects of messengers on inflation forecasts. It is the first forecasting experiment that also causally measures attention to information through interactive design features, enabling novel causal inferences about how both messengers but also inflationary conditions impact attention paid to information. The experiment is also unique in its reflection of credible economic dynamics by using real historical inflation data in a fully controlled within-subject design, where each participant experiences all treatments. The within-subject design most closely relates to McMahon and Rholes (2023), who examine central bank credibility, and offers more flexibility than between-subject randomized control trials (RCTs): For instance, while the thematically closely related RCT by D’Acunto et al. (2022) makes important contributions in highlighting the significance of ingroup effects for central bank communication, my design allows for attributing ingroup effects causally to a single characteristic by varying only the nationality of generic experts in writing. It is also the first to causally estimate the influence of homophily and heterophobia in this context and provides novel causal evidence for policy recommendations. Presenting multiple inflation scenarios further limits the impact of the contemporaneous economic environment, which can affect effect sizes in information provision RCTs (Weber et al., 2025).

Optimal disclosure in coordination games The social value of public information in games with strategic complementarities, first explored by Morris and Shin (2002), has evolved significantly since.² I am the first to consider characteristics of the public signal’s messenger, and to differentiate between heterogeneous agent types based on the alignment with these characteristics in such a model. The heterogeneity in agents determines agents’ information availability and processing. Cornand and Heinemann (2008) discusses partial information availability but my model differs in introducing a belief updating bias and endogenizing signal availability to signal size, extending an idea by Nimark and Pitschner (2019) whereby larger signals are more likely reported. The belief updating bias that I introduce is a resonance weight to expectation formation (Malmendier and Veldkamp, 2022), complementing behavioral distortions such as myopia, or extra discounting of the future, explored by Angeletos and Huo (2021). My model set-up allows me to consider a

Carrell et al., 2010). In healthcare, Alsan et al. (2019) find that Black men are more likely to seek preventative care when interacting with Black male doctors. Similarly, in financial advising, Stolper and Walter (2019) find that homophily significantly increases the likelihood of clients following financial advice.

²They show that public information can disproportionately influence agents’ actions, potentially leading to welfare-reducing overreactions to noisy public signals. Developments since Morris and Shin (2002)’s seminal work include Svensson (2006) arguing that increased precision of public information is likely beneficial for reasonable parameter values, and Amador and Weill (2010) examining a dynamic setting, highlighting varying welfare outcomes.

new communication policy: the delegation of communication. This policy complements the widely explored disclosure policy (e.g., James and Lawler (2011); Angeletos and Lian (2018); Bassetto (2019); Kohlhas (2020, 2022)).

1.2 Three Stylized Facts

This section presents novel motivating observational evidence showing how the messenger’s nationality matters in a monetary union like the Euro area (EA). I derive three stylized facts based on big data of over 8 million multilingual tweets posted between 2016 and 2022: First, the reporting on the messenger (i.e., the ECB president) rises strongly for the ingroup, indicating that the messenger is indeed noticed. Second, information availability—measured as the volume of tweets—is relatively higher for the ingroup. And third, after the ECB’s press conference, Twitter users update their beliefs more strongly when they are in the ingroup.

1.2.1 Twitter Data

To capture real-world patterns of information availability and belief updating, I create a multilingual dataset of publicly available tweets posted between 1 November 2016 and 31 October 2022, using the Academic Twitter API. This time span covers three full years of Mario Draghi’s presidency (2016-2019) until the switchover in ECB presidency on November 1, 2019, and equally three full years of Christine Lagarde’s presidency (2019-2022). To cover the “Big Four” Euro area nations (Germany, France, Italy and Spain), tweets are either in German, French, Italian or Spanish. Tweet language is used to proxy individuals’ nationality.³ To capture general dynamics on Twitter, I also include tweets in English, the language of ECB communication, which unsurprisingly makes up for half of the tweet sample. However, because English language is a poor proxy for nationality, English tweets are excluded for analyses where nationality is key. Collected tweets contain the abbreviation “ecb” or the words “european central bank” or their respective equivalents in any of the mentioned languages.⁴ This gives just over 8 million (8,031,937) tweets after thoroughly cleaning and translating the dataset. Table A.2 provides an overview of the number of tweets by language in the final sample. Details of the cleaning steps and descriptive evidence, including some idiosyncrasies in tweet samples across different languages, are in Appendix A.1.

The Ingroup Individuals are considered to be in the ingroup if their nationality, proxied by tweet language, matches the nationality of the ECB’s president. This means Italians are in the ingroup between November 2016 and October 2019 of the tweet sample, and French

³Appendix A.1.2 demonstrates that tweet language is a valid proxy for nationality. The experiment in this paper will not rely on this assumption.

⁴These terms for selection follow Ehrmann and Wabitsch (2022), except that I exclude “draghi” and “lagarde” because both presidents have/had careers beyond the ECB presidency that are/were of public interest. To avoid picking up tweets about Draghi’s political career in Italy post-ECB, or Lagarde’s IMF presidency pre-ECB, I exclude tweets that mention these two presidents without referring to the ECB. Tweets are only collected if the keywords occur in the original part of the tweet. This includes tweets, retweets, replies or quote retweets.

are in the ingroup between November 2019 and October 2022 respectively.⁵

Why Tweets? Compared to surveys, data from Twitter have several advantages: the sample size is much larger as it captures the full population of relevant tweets, allowing for the depiction of real-world phenomena. The high frequency of tweets, along with the ability to observe the same users multiple times, enables real-time analysis of reactions to central bank communication events and the study of associated belief updates. Unlike surveys, which capture expectations less frequently or only once, Twitter captures spontaneous expressions of opinion from a broad audience, reducing concerns about observer effects (e.g., Hawthorne or experimenter demand effects). While the frequency of tweets is comparable to market data typically used to study financial market reactions to central bank disclosures, Twitter uniquely allows for the study of the wider population rather than just market participants (Ehrmann and Wabitsch, 2022). However, Twitter data also presents some limitations: the evidence focuses on specific real policymakers with various attributes, is influenced by other trends and economic events, and uses imperfect control groups (the outgroup). Although these limitations prevent causal interpretation, the emerging real-world trends highlight key dimensions of messenger effects. In addition, all limitations of Twitter data are addressed with the experiment later in the paper, including answering whether higher tweet volume indicates better reach, whether tweet sentiment translates into inflation expectations, and what drives these findings.

1.2.2 Stylized Fact 1: Higher Reporting on Ingroup Policymakers

The first stylized fact that emerges from analyzing tweets is that the reporting on the main ECB policymakers varies strongly by language: ECB presidents are mentioned much more in tweets if they are in a language's ingroup than outgroup. This fact is important to establish that ECB presidents are indeed talked about, and their nationalities play a role for how information gets reported. Thereby supporting that any further identified differences between languages could indeed be due to being in- and outgroup with the ECB policymaker.

To highlight the different focus on policymakers, Figure 1.1 shows how frequently ECB presidents are talked about in the various language samples. Recall that tweet samples are only selected based on ECB-related key words, not based on ECB presidents. Notably, while “draghi” ranks among the top 10 most frequent words in the Italian sample, “lagarde” only ranks around the top 50. Similarly, in the French sample, “lagarde” ranks among the top 10, while “draghi” is around the 60th most frequent word. In the English sample, the two policymakers are within 6 ranks, and similarly in the German and Spanish sample, the rankings of the two policymakers are much closer.

Thus, while across languages ECB presidents are talked about in connection with the ECB, the reporting on these policymakers is heterogeneous, with a stronger focus on ingroup policymakers. This highlights that the nationality of ECB presidents alone might be enough to impact the availability and content of information, thereby impacting how effective central bank communication can be.

⁵Note that the experiment in this paper will abstract from specific policymakers, defining the ingroup as participants who share the same nationality as the generic messengers.

Figure 1.1. Reporting on Policymakers by Language



Notes: Word clouds show the 100 most frequent words in English (top), in Italian (bottom left) and French (bottom right) between 2016 and 2022. Word size indicates word frequency. “Christine”, “Lagarde”, “Mario” and “Draghi” are highlighted in green for better visibility. Word clouds are based on cleaned, translated and lemmatized tweets. Stopwords are removed. No ngrams are used. Word clouds in German and Spanish are found in Figure A.2 in Appendix A.1.

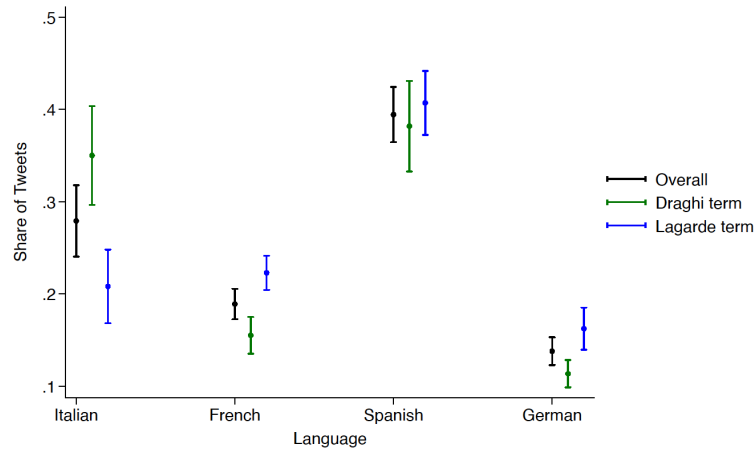
1.2.3 Stylized Fact 2: Higher Information Availability

The second stylized fact that emerges from the data is that information availability about the ECB increases for the ingroup, indicating a better chance of reaching a wider audience.

To show this, I analyse tweet volume by language and find that tweet volume about the ECB is relatively higher when a nationality is in the in- rather than the outgroup. Figure 1.2 plots the shares of tweets by language in six-week intervals corresponding to the ECB's press conference cycle. Shares are shown for the entire duration of the sample (in black) and split into the sample years covering Draghi's presidency (in green) and Lagarde's presidency (in blue). Using shares of 6-week periods is beneficial to reduce concerns about time trends driving the results (e.g., avoiding a rise in Twitter volume over time because of user adoption being wrongfully associated with Lagarde-related tweet behavior), but also to avoid a few events driving the results. Overall, the Spanish tweet the most (well over a third of tweets), followed by the Italians (just under 30%). The French and Germans are much less active on Twitter for this topic. Comparing the differences in tweet shares between Draghi's term and Lagarde's term, Italians tweet substantially more during Draghi's presidency, while the opposite is true for the French: their share in tweets rises by around 50% during Lagarde's term. The Spanish and German samples serve as a type of control group.⁶

⁶Indeed, the Spanish tweet volume does not change much between presidencies. However, German tweets does increase slightly during Lagarde’s term. While this may be attributed to Lagarde, this might also be driven by Isabel Schnabel—who is very active on Twitter—joining the ECB Executive Board just two months after Lagarde. Further, this might even merely be a composition effect, assuming that (for whatever reason) German Twitter was growing relatively more than the average during this time. Appendix A.1 discusses some of the most important observed trends by language. In sum, while the patterns strongly suggest that being in the ingroup increases information availability on Twitter, the presence of confounding factors prevents any

Figure 1.2. Share of Tweets by Language



Notes: The plot shows the share of tweets by language per 6-week cycle between ECB Press Conferences (with 95%-CI). The sample considered includes tweets about the ECB in German, Spanish, French and Italian between 2016 and 2022. The color of dots and error bars indicate the time horizon considered for the share of a language's tweets: black uses the entire sample, while green (blue) uses the sub-sample during Mario Draghi's (Christine Lagarde's) presidency.

A simple regression reveals a significant positive ingroup effect on information availability via Twitter: for the ingroup, information availability increases by 10.5 percentage points. These ingroup effects are positive across all levels of monetary expertise, with the strongest impact observed among users who are neither strict experts nor complete non-experts, where the (non-)expert classification follows the benchmark in Ehrmann and Wabitsch (2022) (see Table A.4).

I confirm the information availability patterns found on Twitter with traditional media, using the volume of printed newspaper articles about the ECB. The ECB president's nationality significantly increases the share of their corresponding national newspaper volume by 6.1 percentage points (or by roughly 272 articles). Appendix A.1.4 shows the full newspaper analysis.

Later in this paper, I show that information is not only more available, but individuals are also better reached by ingroup policymakers (see Section 1.3.6 and Appendix A.2.4).⁷

In sum, information availability about the ECB increases for the ingroup, making central bank communication more effective at reaching this group. Next, I will assess whether the received information is also more effective at shifting beliefs of the ingroup.

1.2.4 Stylized Fact 3: Stronger Belief Updating

The final key takeaway from Twitter is on information processing. Following central bank communication, the ingroup updates their beliefs more strongly. For this, I analyze the content of tweets using common natural language processing (NLP) methods.

causal interpretation, as often is the drawback of observational data.

⁷This finding is based on a survey about receiving news about central bank policymakers that is embedded in my experiment.

The Signal The most prominent communication event of the ECB is its press conference, which takes place every six weeks.⁸ It is particularly suitable to study, as it is the ECB’s main communication event where any rate changes are announced, a statement on the economy is given, and questions from the press are taken. It is well-anticipated by observers and communication is carefully designed by the central bank. The final sample contains 48 press conferences, where the first took place on December 8, 2016 and the last on October 27, 2022.

The Beliefs I assess the sentiment of tweet content as a broad indicator of individuals’ beliefs. There are various NLP approaches to estimating text sentiment, each with its own advantages and disadvantages. While the results are robust to more complex NLP methods, I show results based on a dictionary-based sentiment indicator due to its transparency and simplicity. The dictionary-based method provides a continuous estimate of how positive or negative a text snippet is.⁹ The distribution of sentiment is highly similar across all languages, precluding any idiosyncratic differences in tweet sentiment between them. This, alongside details of measuring sentiment, is shown in detail in Appendix A.1.

Relative Belief Updates To assess individual belief updates, I compare the sentiment of tweets before and after ECB press conferences. Belief updates are calculated as the absolute change in sentiment between a user i ’s last tweet during the ECB’s quiet period—7 days before the conference— $\text{Sentiment}_{i,t-1}$ and the sentiment of the same user’s first post-conference tweet posted within 24 hours of the conference $\text{Sentiment}_{i,t}$, relative to the average update around each press conference to account for differences in signals.¹⁰ The absolute difference captures the magnitude of belief change, regardless of direction, and subtracting the average belief update of all individuals at conference j controls for variations in information content or timing, focusing on deviations from the common trend. This approach captures responses to the press conference and avoids picking up other ECB communications due to the quiet period’s restricted communication. Figure A.4 in the Appendix shows the distribution of sentiment before and after press conferences.

Figure 1.3 plots these relative belief updates by nationality for the entire duration of the sample (in black) and split into press conferences led by president Draghi (in green) and president Lagarde (in blue). The following conclusions can be drawn: First, Italians generally update more strongly than others. Second, Italians and French update their beliefs more strongly when they are in the ingroup—the Italians under Draghi and the French under Lagarde—rather than the outgroup. Again, the Spanish and German sample act as a control.¹¹ Based on these four nations, the overall estimated ingroup effect on belief updating estimated by a simple OLS is 0.014***.

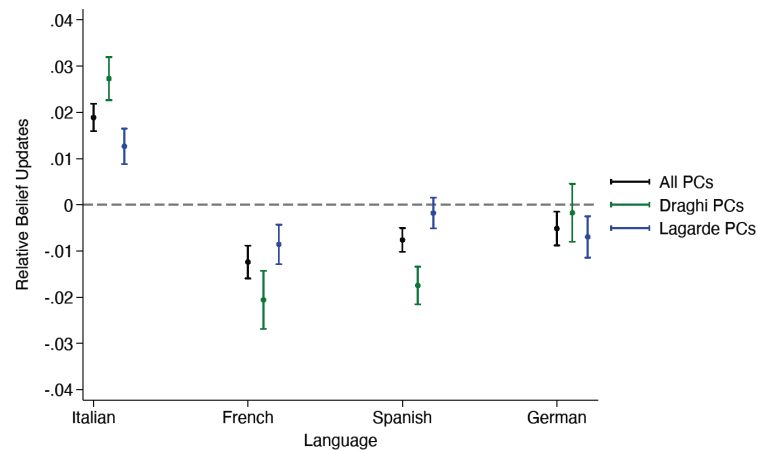
⁸The monetary policy announcement changed from 13:45 to 14:15 CET after July 21, 2022. Similarly, the starting time of the press conference shifted from 14:30 to 14:45 CET (see [here](#)).

⁹I use the polarity measure integrated in the Python package TextBlob, which is based on the Princeton University’s WordNet lexicon (Loria, 2018).

¹⁰Formally : $\text{Relative Update}_{ij} = |\text{Sentiment}_{i,t} - \text{Sentiment}_{i,t-1}| - \frac{1}{N_j} \sum_{k=1}^{N_j} |\text{Sentiment}_{k,t} - \text{Sentiment}_{k,t-1}|$

¹¹While the Germans show no difference in updating behavior between presidents, the Spanish are updating more strongly during Lagarde’s presidency. Instead of this representing a Lagarde effect, however, this might instead be driven by the Spanish de Guindos. As the vice president of the

Figure 1.3. Belief Updates by Tweet Language



Notes: Plot shows the absolute size of belief updates relative to the average during a given ECB Press Conference (with 95%-CI). The sample considered includes tweets about the ECB in German, Spanish, French and Italian between 2016 and 2022. The color of dots and error bars indicate the time horizon considered for the share of a language's tweets: black uses the entire sample, while green (blue) uses the sub-sample during Mario Draghi's (Christine Lagarde's) presidency.

Finding that the ingroup exhibits larger absolute changes in sentiment suggests that they update their beliefs more strongly in response to central bank communication. But do these larger belief updates indicate that the ingroup shifts their beliefs more towards the information communicated by the central bank? While the presented descriptive evidence does not allow for a definitive conclusion, a detailed analysis accounting for the actual communicated signal in Appendix A.1.5 offers further insights: While all individuals' beliefs are responsive to novel information from the central bank, ingroup agents are more susceptible than outgroup agents. Ingroup agents rely less on their existing beliefs when updating beliefs and instead place more weight on new information from the central bank. They use new information over twice as much as outgroup agents, while relying on existing beliefs about half as much (see Table A.7). These ingroup effects persist regardless of individual monetary expertise, though they are slightly more pronounced among non-experts than experts as visualized by Figure A.5. Overall, the heightened responsiveness implies that communication is more effective within the ingroup.

1.2.5 Discussion

The evidence I provide demonstrates that nationality of the messenger matters in a multi-national monetary union like the Euro area. When individuals are in the ingroup with the messenger, information is more available and beliefs respond more strongly. These ingroup effects are present for both experts and non-experts. Thus, it is not solely journalists reporting differently about the ECB that drives these results. While this evidence provides valuable insights into real-world phenomena, it is affected by the aforementioned data limitations and cannot inform on the underlying mechanisms. To address this, and to establish the causal effect of the messenger on inflation expectations, I design an experiment.

ECB, he is also present at press conferences and his term time largely coincides with Lagarde's term.

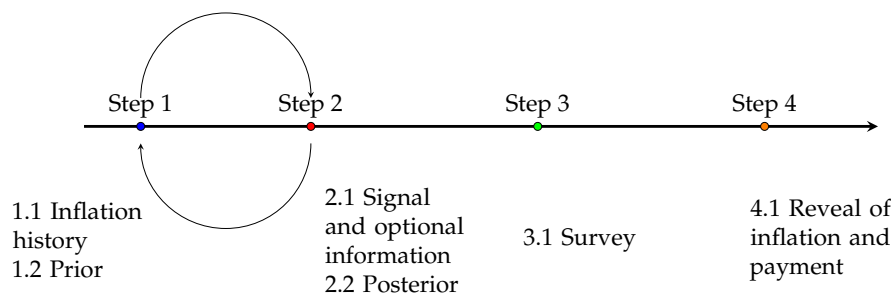
1.3 Inflation Forecasting Experiment

This section describes the individual-choice inflation forecasting experiment and its findings. In this pre-registered experiment, participants in Germany, Spain, France and Italy forecast inflation in six different inflation scenarios, being given information by six different messengers (the treatments). The experiment is explicitly designed to provide causal effects of messenger characteristics on belief updating about future inflation, the mechanism that drives differences in belief updating, and messenger-driven attention to information. The design allows me to identify the size of the nationality-based ingroup effect on belief updating, and whether such effects are rooted in homophily rather than heterophobia. Finally, findings inform two potential policy recommendations: First, can the ECB communicate more effectively by strategically communicating using different board members more prominently? And second, in a similar vein, can beneficial ingroup effects be reaped by communicating more through the network of the Eurosystem’s national central banks instead of centralized communication via the ECB?

1.3.1 Experimental Design

Methodologically, this inflation forecasting experiment is individual-choice and follows primarily a within-subject design. It is incentivized, fully controls the environment in which participants forecast, and most closely relates to McMahon and Rholes (2023) who study credibility varying the forecast performance of central banks. In my experiment, two decisions that participants make are key: forecasting inflation and requesting additional information (or not). After participants complete all forecasting tasks, they are surveyed on their perceived ability of messengers, on their trust in institutions and policymakers, and on their news consumption of various policymakers and monetary policy. This allows me to draw conclusions on ingroup effects for inflation expectations, the driving mechanism, whether the ingroup is reached more and causally pays more attention to information. The timeline of the experiment is shown in Figure 1.4 and each aspect is explained in detail below.¹²

Figure 1.4. Timeline of Experiment



Notes: The timeline goes from left to right. Steps 1-2 repeat for all 6 forecasting tasks before going on to step 3 and 4.

Inflation Forecasting Tasks At the core of this experiment are inflation forecasting tasks, which measure how participants update their inflation expectations following information

¹²The experimental interface design facilitates task comprehension. It incorporates interactive tools and visuals, and is coded in oTree (Chen et al., 2016). See Appendix A.3 for screenshots of the experimental interface.

provided by randomly varying messengers. Updates of inflation expectations are later evaluated through a standard Bayesian belief updating framework. What this means is that posterior beliefs should be determined by the value and the precision of both the prior and the signal. Thus, it is important that the experimental design not only provides values of the signals and their precision, but also collects data on participants' priors and their precision.

Each forecasting task follows the same structure: First, participants observe 10 periods of inflation history before submitting their own inflation point and range forecasts for the next period (period 11).¹³ This point forecast represents their prior, and the corresponding range forecast their inverted prior precision, which indicates participants' uncertainty associated with their prior. Participants will then receive a professional forecast for period 11 (i.e., the signal). This signal is provided as text but also appears in the interactive chart (see Figure A.18). The forecast is linked to randomly varying messengers, and comes together with the given messenger's forecast history of the previous 10 periods. This forecast history allows participants to assess the signal precision via the messenger's historical forecast error (comparing the historical forecasts to realized inflation).¹⁴ While this requires the assumption that participants perceive historical forecast errors as the signal noise—or at least do not change their interpretation of provided signal noise between treatments—, this reflects how forecast quality and credibility is typically assessed in the real world. The historical forecast performance is visually displayed in a chart, with the exact values appearing when participants hover over data points. I refer to each combination of inflation and forecast values as a scenario. To preclude any differences in the six used scenarios from impacting effects, scenarios are randomized between messenger treatments. Finally, participants are given the chance to update their inflation forecast (point + range), i.e., they provide their posterior.¹⁵

Messenger Treatments The forecasting tasks are designed with ten messenger treatments—out of which each participant faces six—to precisely test four hypotheses (H1-H4). Table 1.1 gives an overview of all treatments and their alignment with the hypotheses.¹⁶ Treatments are grouped into three categories: *Generic Expert Treatments*, *ECB Expert Treatments*, and *Institutional Expert Treatments*.

1. *Generic Expert Treatments*: These quantify the pure ingroup effect by assessing how individuals use information from generic experts distinguished only by nationality. Participants encounter two messengers: one by an expert of their nationality (*generic ingroup expert*)

¹³Range forecasts are participants' expected upper and lower bounds of inflation. Input values are restricted to ensure that participants' point forecasts are within their provided range. The experimental software interactively displays the forecasts and inflation history of a given sequence. For an example of this, see Figure A.17 in Appendix A.3.

¹⁴10 data points of forecasted and realized inflation allow participants to assess the precision of the signal fairly well. Signal precision is proxied by the inverse average forecasting error of the 10 periods. Table A.8 gives an overview of the values used for signal precision.

¹⁵Figures A.17 and A.18 in Appendix A.3 show the interface for these forecasting tasks. Instructions emphasize that participants may either use completely new values for their updated forecasts, some new values, or stick entirely to the values of their initial forecast, of which they are reminded (see Figures A.14 and A.15 in Appendix A.3 for the full instructions).

¹⁶Section 1.3.4 discusses the hypotheses and findings and Appendix A.2 provides more details on how treatments correspond to hypotheses.

Table 1.1. Overview of Messenger Treatments

Treatment	Hypotheses	Messenger
1	H1	Expert from France
2	H1	Expert from Italy
3	H1	Expert from Germany
4	H1	Expert from Spain
5	H2, H3	Expert from France who represents the ECB
6	H2, H3	Expert from Italy who represents the ECB
7	H2, H3	Expert from Germany who represents the ECB
8	H2, H3	Expert from Spain represents the ECB
9	H3, H4	Expert who represents the European Central Bank (ECB)
10	H4	Expert who represents the National Central Bank (NCB)

Notes: The order of treatments is randomized at the participant-level. Additionally, the forecasting sequences are randomly allocated to treatments for each participant. In the experiment, font color is uniformly black and font colors displayed here simply aim to provide a better overview of the treatments to the reader. H1-H4 refer to the hypotheses that the treatments will be testing, which are explained in more detail in Section 1.3.4. For treatments 1–4, and 5–8 participants face only two out of the four treatments: the in- and one outgroup messenger (where the outgroup messengers' nationality is held constant within a participant). All participants face treatments 9 and 10.

and one by an expert of a randomly selected nationality out of the three other nationalities (*generic outgroup expert*).

2. *ECB Expert Treatments*: These add institutional context by introducing messengers representing the ECB with specified nationalities. As for the generic expert treatments, one ECB expert matches the participant's nationality (*ECB ingroup expert*), while the other ECB expert does not (*ECB outgroup expert*). The latter messenger's nationality matches the nationality of the generic outgroup expert.

3. *Institutional Expert Treatments*: These involve generic experts without specified nationalities representing either the ECB or the participant's national central bank (NCB).

To avoid fatigue, each participant faces only six out of the ten treatments—two from each group. Treatment order and inflation scenarios are randomized at the participant-level to ensure valid comparisons.

Inflation and Forecast Data Data for the forecasting tasks are based on actual annual inflation realizations in the Euro area between 1990Q1 and 2023Q1 (considered at quarterly frequency), and real-world quarterly forecasts of Euro area annual inflation provided by the ECB Survey of Professional Forecasters (SPF).¹⁷ Using real data ensures that economic dynamics are credible. To emphasize this, participants are explicitly told that the dynamics of inflation and its forecasts always follow the real Euro area economy and that forecasts are real historical expert forecasts. The large amount of data points (97 quarters) make it unlikely that participants would recognize a historical pattern, and successfully predict the next period. Six sequences are randomly selected—one for each of the 6 forecasting tasks that participants complete. Appendix A.2 shows the quarterly data series considered for

¹⁷The SPF is a quarterly survey of expectations for the rates of inflation and other variables. Respondents of the SPF are experts affiliated with financial or non-financial institutions based within the European Union. More information on the SPF can be found [here](#).

the experiment (Figure A.6) and the forecasted and realized inflation for selected sequences (Figure A.7), as well as more details on the forecasting precision averages, realized inflation averages, and standard deviations of realized inflation for 6 randomly selected sequences (Table A.8).

Incentivizing Forecasts Participants are incentivized to minimize forecast errors. Applying common incentive structures from the experimental literature ensures decisions are theory-consistent and of high quality (Charness et al., 2016). Similar to McMahon and Rholes (2023), I incentivize priors, posteriors, and their respective precisions, using methods from the learning-to-forecast literature (e.g., Rholes and Petersen (2021), Mokhtarzadeh and Petersen (2021)).

For *point forecasts*, participants receive a bonus payment based on their forecasting score, which decreases with the absolute forecast error. A perfect forecast yields the maximum score, and the score diminishes as the forecast error increases, reducing by a consistent proportion for each percentage point of error.

For *range forecasts*, participants earn a payoff only if the realized inflation falls within their specified forecast range. The payoff decreases as the width of the forecast range increases. This scoring rule incentivizes participants to keep their forecast ranges as narrow as possible while still covering inflation values they consider probable.

More details on bonus payments, including how incentivizing bonus payments are calculated and randomly selected are shown in Appendix A.2.

Attention to Information In addition to the passively endowed forecast of a messenger (i.e., the signal), participants may actively choose more information from a given messenger. This provides a measure of attention to central bank communication. Participants are offered the opportunity to read more about the messenger’s economic analyses underlying the given forecast. These analyses are summaries of historical ECB communication that were released roughly at the same time as the forecasts from the SPF.¹⁸ This aims to mimic real central bank communication while ensuring that the numerical signal effectively summarizes the accompanying qualitative information. In the experiment, there are three “Read More” buttons, which participants have to click on to reveal the optional information. This measures attention by tracking participants revealing information piecewise, similar to Bartoš et al. (2016) who count clicks on “learn more” buttons to measure acquired information about job applicants’ résumés. One can think of the pieces of information as a flow of information, for which participants decide when to “stop”, leaving them with their demanded level of acquired information. An overview of all additional information pieces and their corresponding inflation scenario is given in Table A.9 in Appendix A.2. A widely used alternative measure of attention is the amount of time spent on making a decision. However, this proxy of attention suffers from substantial drawbacks, such as being unable to differentiate causes of short decision times (e.g., fast thinking vs. low attention or unrelated interruptions).

¹⁸This ensures the given additional information links to the corresponding forecast scenario.

Post-Experimental Survey After completing the forecasting tasks, participants fill out a survey aimed at understanding the mechanisms behind their decisions.¹⁹

Participants first rated the *perceived ability* of the treatment messengers to forecast inflation and provide economic analyses on a 7-point Likert scale. This measure identifies any prejudice or experience that might have influenced their perceived signal precision of the messengers. Accounting for these perceptions allows me to distinguish between perceived signal quality, and remaining discrimination beyond perceived signal quality, which participants are not aware of and may be viewed as a bias.

To account for pre-existing perceptions of institutions and policymakers that could influence decision-making, participants report their *trust* and *exposure* (the extent to which they know of and follow news) regarding monetary institutions (the ECB and participants' respective NCB) and representative real-world policymakers from messenger treatments. The representative policymakers are NCB governors and ECB board members of the same four nationalities at the time of the experiment, Table A.10 provides an overview of chosen representative policymakers.

Finally, they were asked about their *monetary policy expertise* and were tested on their attention during the experiment.

1.3.2 Experimental Sample

The experiment is run in four different countries (the big four nations in the EA) to ensure that any identified effects that may be culturally driven are fairly representative of the EA.²⁰ While the random sample is not perfectly representative of the populations in Germany, Spain, Italy or France, it captures the populations fairly well regarding employment status, education and income. Although participants are slightly younger than the population average, the sample is much more diverse compared to common samples in lab experiments, as these commonly include students only. Appendix A.2.3 gives a detailed overview of participants in the sample. Participants' nationality is required to also coincide with the country of residence to preclude the inclusion of participants with mixed or potentially weaker cultural identification due to living abroad. In total, 400 participants (100 per nationality) do the experiment online via Prolific. One outlier is removed from the Spanish sample, leaving the final sample at 399.²¹ A power analysis can be found in Appendix A.2. To ensure participants understand how the experiment and their incentives work, they must pass a comprehension quiz on the instructions to be allowed to proceed.²² The vast majority of data is collected on October 31, with a few responses being added within 48

¹⁹All survey questions are answered after the forecasting tasks to avoid priming but before revealing the realized values, preventing task performance from influencing responses about similar policymakers. Details of the survey as well as the entire survey are shown in Figures A.20-A.25 in Appendix A.3.

²⁰This is a between-subject element of the experiment.

²¹The outlier is dropped due to forecasted inflation values being consistently above 20, going up as high as 70, which indicates complete disregard of how forecasts are incentivized, and thus casts doubt as to whether instructions have been carefully read.

²²Overall, 553 participants start the experiment, out of which 153 drop out because they either fail the quiz on instructions, voluntarily return the experiment before passing the quiz, or time out (only very few do so).

hours. On average, participants earned £3.58 in bonus payments and thus £7.58 overall.²³

1.3.3 Estimating Belief Updating

Framework to Assess Belief Updating I use a standard Bayesian belief updating framework to assess how agents update their inflation expectations following central bank communication.²⁴ The posterior belief is a weighted average of the prior and the signal, based on their relative precision (e.g., Veldkamp (2011)). Appendix A.2.2 provides more explanations.

Experimental evidence suggests that people are not perfectly Bayesian and tend to under-infer from signals.²⁵ Whether this is the case here, I will now test empirically.

Empirical Estimation This subsection shows the empirical estimates of how participants use their priors and signals. According to the theoretical framework, posterior beliefs are a weighted combination of prior beliefs and received signals, with weights determined by their relative precision.²⁶ I test this empirically using the following regression:

$$Posterior_i = \gamma \underbrace{\left(\frac{\alpha_i}{\alpha_i + \beta} A_i \right)}_{\text{weighted Prior}} + \delta \underbrace{\left(\frac{\beta}{\alpha_i + \beta} B \right)}_{\text{weighted Signal}} + \epsilon_i, \quad (1.1)$$

where the coefficients γ and δ indicate whether agents act Bayesian. If $\gamma = \delta = 1$, agents update their beliefs according to Bayes' rule. The parameter γ captures the influence of the prior: $\gamma > 1$ implies over-reliance on priors, while $\gamma < 1$ indicates under-use (base-rate neglect). Similarly, δ reflects responsiveness to new information: $\delta < 1$ means under-use of signals, and $\delta > 1$ implies over-use.

Estimating Equation 1.1 by OLS without a constant yields $\gamma = 1.21$ and $\delta = 0.90$. This indicates that compared to a Bayesian benchmark, the signal is substantially under-used, regardless of nationality (see Figure 1.5 and Table A.11 in the Appendix). Participants under-use the signal by about 10 percentage points on average, aligning with findings in

²³Participants earn £4 for completing the experiment and may earn up to £6 in bonus payment, depending on the accuracy of their point and range forecasts.

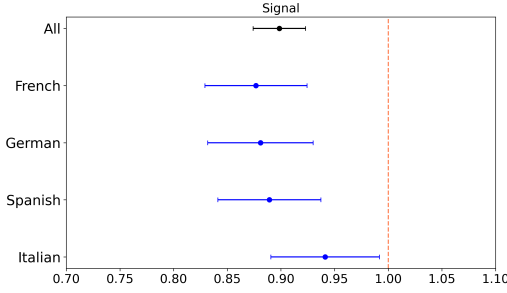
²⁴Inflation data and professional forecasts may not perfectly align with the theory, which is reflected in the interpretation of results.

²⁵Benjamin (2019) provides an overview of biases in belief updating. This under-inference may be due to factors like extreme-belief aversion (Benjamin et al. (2016)) or conservatism bias, where beliefs are updated too slowly, as discussed by Phillips and Edwards (1966).

²⁶As outlined in Section A.2.2, the Bayesian benchmark assumes inflation is normally distributed and the signal is independently and identically distributed (i.i.d.), requiring forecast errors to be serially uncorrelated. However, historical quarterly EA inflation is non-normally distributed, and inflation forecasts show slight autocorrelation, invalidating the signal's i.i.d. assumption. Thus, I refer to a Bayesian *benchmark* that assumes i.i.d. signals and normally distributed inflation. Deviations from this benchmark do not necessarily imply non-Bayesian behavior; it serves as an imperfect reference point. Importantly, I focus on how participants' updates differ between messenger treatments, which are equally affected by these data limitations.

existing literature.²⁷ This observed signal under-use suggests that increasing signal use is beneficial for closing the gap with a Bayesian benchmark. Can certain messengers achieve this?

Figure 1.5. Signal Use Across Treatments by Nationality



Notes: The plot shows the regression coefficients of signal use with 95% confidence intervals. Results for the entire sample is shown in black and results for samples split by nationality are shown in blue. The dashed vertical line at 1 indicates the Bayesian benchmark for signal use.

To test how different messengers affect signal use, I expand Equation (1.1):

$$Posterior_{ij} = \gamma \left(\frac{\alpha_{ij}}{\alpha_{ij} + \beta_j} A_{ij} \right) + \sum_{j=1}^J \delta_j T_j \left(\frac{\beta_j}{\alpha_{ij} + \beta_j} B_j \right) + \epsilon_{ij}, \quad (1.2)$$

where T_j is a dummy variable that equals 1 for each messenger treatment $j \in J = \{1, 2, \dots, 6\}$. For each treatment j , I estimate a separate δ_j coefficient to test hypotheses.

One caveat is that, due to the limited sample size, the randomized matching of messenger treatments with task order and inflation scenarios is imperfectly balanced.²⁸ To ensure that I identify variation in messenger treatments rather than inflation scenarios, I control for inflation scenarios and the order of messenger treatments in the analysis. While including these controls properly identifies the treatment effects, the Bayesian interpretation of the absolute size of coefficients is no longer valid. Therefore, I focus on the relative signal uptake between treatments to test my hypotheses.

1.3.4 Causal Findings on Updating Inflation Expectations

This subsection presents how different messengers impact belief updating, providing the pure ingroup effect and causal evidence for two concrete policy examples of delegating communication, and explores the mechanism behind the observed ingroup effects.

To obtain a clean causal nationality-based ingroup effect on updating inflation expectations that addresses a core research question of this paper, I test the first hypothesis:

²⁷For example, Benjamin (2019) find a δ coefficient of 0.86 in a meta-analysis of incentivized tasks with a single signal.

²⁸With a sufficiently large sample, randomization of inflation sequences and messenger treatment order (as described in Section 1.3.1) should balance out any effects. However, perfect balancing is not achieved here due to sample size limitations.

Hypothesis 1: *There exists a pure ingroup effect to updating inflation expectations, when signals come from experts of varying nationality.*

To test this, I take the *Generic Expert* treatments (1-4), and compare signal use between in- and outgroup experts. These experts are generic with the only information known about them being their nationality, which is made salient to participants (i.e., “an expert from Germany/France/Italy/Spain”). Except for the attached nationality, messaging experts are identical, allowing for a clean identification of a pure causal effect by avoiding having to account for any features related to a real person or institutions.

I find support for Hypothesis 1: National ingroup effects are significant in central bank communication, with ingroup members showing stronger belief updates. The size of this causal nationality-driven ingroup effect is 0.052*** (see column (3) in Table 1.2).²⁹ Meaning, while participants under-use signals of both in- and outgroup experts, they use the signal of the ingroup expert roughly 5 percentage points more, when forming their posterior beliefs about inflation. Putting this effect size in perspective, the ingroup effect is akin to closing half of the sample average gap to the Bayesian benchmark (as shown in Section 1.3.3). Thus, this ingroup effect to updating inflation expectations is sizable and positively impacts the effectiveness of a signal.

Since central banks do not communicate without institutional context, it is relevant to consider how institutional reputation may affect policymakers’ perception and credibility. Does the ingroup effect persist within the institutional context?

Hypothesis 2: *There exists an ingroup effect when signals come from ECB experts of varying nationality.*

This expands the pure ingroup effect from hypothesis 1 by adding the ECB as the institution of the expert. I compare signal use of messenger treatments 5-8 (ingroup vs. outgroup ECB experts).

I find support for Hypothesis 2, albeit somewhat weak: The ingroup effect diminishes to roughly a quarter of the average gap to the Bayesian benchmark (0.028*) once the ECB is introduced as the institution of the expert (see Table 1.2). This effect is the causal ingroup effect within the ECB context but adding the institutional dimension may introduce noise. Two sources of noise seem particularly likely: First, the institutional information may trigger associations beyond nationality that may cause other in- or outgroup effects, weakening the nationality treatment (e.g., along the dimensions of level of education, field, perceived elitism, geographic location, etc.). Stolper and Walter (2019) show that various attributes can simultaneously cause ingroup effects. Second, even though the effect holds the institutional context constant across treatments, it is now plausible that the combination of nationality and ECB triggers associations with real policymakers. Participants might then distort the provided signal precision by what they know about the forecast ability of these associated policymakers, as suggested by Esponda et al. (2023). Besides messenger ability, also trust could then play a role, which is an important factor in the estimates as I will show in Section 1.3.5. Considering these caveats, my diminished results compare somewhat

²⁹This effect is based on the estimation that controls for inflation scenario, order of treatments and individual fixed effects.

Table 1.2. Main Results from Experiment (H1, H2, H3, H4)

	(1)	(2)	(3)
<i>Pure Ingroup Effect (H1):</i>	0.064** (0.029)	0.047*** (0.017)	0.052*** (0.017)
R-squared	0.960	0.986	0.994
N	795	795	795
<i>Ingroup Effect for ECB Experts (H2):</i>	-0.014 (0.029)	0.022 (0.018)	0.028* (0.017)
R-squared	0.960	0.985	0.994
N	795	795	795
<i>Homophily - Ingroup ECB Expert vs. Generic ECB Expert (H3):</i>	0.006 (0.029)	0.034* (0.018)	0.035** (0.018)
R-squared	0.961	0.985	0.993
N	794	794	794
<i>Heterophobia - Outgroup ECB Expert vs. Generic ECB Expert (H3):</i>	0.020 (0.027)	0.014 (0.017)	0.013 (0.017)
R-squared	0.965	0.986	0.994
N	795	795	795
<i>NCB vs ECB: Institutions Effect (H4):</i>	0.027 (0.028)	0.031* (0.017)	0.034** (0.017)
R-squared	0.965	0.987	0.994
N	795	795	795
Inflation Scenario		✓	✓
Individual-FE			✓

Notes: Effects shown here result from comparing coefficients of interest of the same regression. *Inflation Scenario* refers to controlling for the underlying data of forecasting tasks and the order in which they appeared. *N* refers to the number of observations (forecasting tasks) in the regression. *N* may slightly vary across treatments due to instances of infinite prior precision. Column (3) represents the main results. Stars correspond to the following p-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

to D’Acunto et al. (2022) who do not find significant results for inflation expectations when varying messengers of the same institutions. However, their study differs from my experiment in an important aspect: The use of various real policymakers as messengers, which means multiple messenger characteristics vary simultaneously between treatments. As my experimental design only varies a single characteristic at a time, any potential noise from other characteristics is avoided, which could further explain why I find significant effects for inflation expectations, while D’Acunto et al. (2022) do not.

An alternative interpretation of the diminished effect is that the ECB as an institution carries enough weight to dwarf the relevance of its policymakers’ nationalities. Given the ingroup’s greater endogenous exposure to news from relevant policymakers (as shown by survey data in Section 1.3.7), even a modest ingroup effect on information processing may result in significant real-world differences in belief updating, as indicated by the belief updating patterns on Twitter.

The positive messenger effects on information availability and processing within the ECB context should be reflected in French households' inflation expectations, given the French nationality of the current ECB president. I find corroborating correlations of this pattern using cross-country representative survey data from the ECB's Consumer Expectations Survey (CES) between the start of the survey in 2020 and up until 2024. Compared to other nations, the French reported inflation expectations that are most aligned with ECB forecasts (see Figure A.12).³⁰ The detailed analysis can be found in Section A.2.5 in the Appendix.

What is the root cause of these ingroup effects within the ECB context? Differential signal use between two ECB experts may stem from a relative over-use of the ingroup expert's signal (homophily), a relative under-use of the outgroup expert's signal (heterophobia), or a combination of both. To understand which one it is, I now test:

Hypothesis 3: *Higher signal use for ingroup experts (homophily), rather than lower signal use for outgroup experts (heterophobia), drives the ingroup effect within the ECB context.*

I take advantage of having ECB expert messenger treatments with and without nationalities attached. Homophily compares participants' updates to the signal from ECB ingroup expert vs. the generic ECB expert, and heterophobia compares the signal use of ECB outgroup experts vs. generic ECB expert (treatments 5-8 vs. 9).

I find support for Hypothesis 3. Participants update significantly more towards the ingroup ECB expert than the generic ECB expert (0.035**), while they do not update significantly less to the outgroup ECB expert compared to the generic ECB expert (0.013). Thereby, ECB ingroup effects are driven by homophily, not heterophobia (see Table 1.2).

Instead of varying messengers within an institution, could varying institutions return similarly affect belief updating? If NCBs are seen as an "ingroup institution", while the ECB is seen as an "outgroup institution", the ingroup effect might translate to institutions (without attaching an explicit nationality to messengers). The final hypothesis tests:

Hypothesis 4: *There exists an ingroup effect when signals come from national central banks (NCBs) rather than the ECB.*

I find support for Hypothesis 4 by comparing the signal use between treatments 9 and 10 across all participants: Participants update more towards the NCB signal over the ECB signal (0.034**)—a third of the average gap to the Bayesian benchmark). Effects are smaller than the pure ingroup effect but larger than the ECB ingroup effect. A possible explanation for the former is that institutions are not strictly viewed as "in- vs. outgroup institutions"; after all, the ECB is the supranational institution. Another explanation is that institutional reputation plays a big role, which is likely not constant across institutions. Finally, adding institutions to messenger treatments potentially adds noise, as discussed above for Hypothesis 2. However, for the NCBs in the experiment, the found effect is causal and meaningful.

³⁰ECB forecasts are taken from the ECB's Macroeconomic Projection Database (MPD), which is published four times a year and contains information on the outlook for the EA and contributes to the ECB Governing Council's assessment of economic developments and risks to price stability.

This positive effect on belief updating in response to NCB communication suggests that delegating communication to national banks within the Eurosystem could be more effective than centralizing it through the ECB. Yet, this might come at costs and risks such as the cacophony of voices (Blinder, 2007). It is vital that the central bank carefully assesses the number of speeches given, and ensures focused and cohesive messaging (Do Hwang et al., 2021; Tutino, 2016). Overall, the benefits of delegating communication to NCBs highlighted here should be viewed as an upper bound, assuming that the cacophony of voices is avoided by ensuring identical messaging.

1.3.5 Mechanism: Perceived Quality and Trust

How conscious are participants of their discrimination between messengers? Are ingroup effects fully reflecting subjective perceptions of messenger ability (i.e., signal quality), or trust in real-world policymakers and institutions? Indeed, much of the effects correlate with these factors.

Observed differences in information use may reflect perceived signal quality, beyond the provided signal precision in the experiment. Participants might perceive their ingroup messengers as more able (i.e., perceiving their signals as more precise) based on subjective experience, disregarding the provided forecast history. To understand the extent to which participants' information use corresponds to the subjectively perceived information quality, I interact messenger-dependent signal use with the self-reported perceived ability of a messenger to forecast inflation and provide economic analyses. The findings are in column (3) of Table A.12 in Appendix A.2.3. Approximately 80% of the pure ingroup effect can be attributed to reported perceived ability. Since perceived ability is reported ex-post, it may also reflect participants' justification for their signal use beyond their genuine perceptions of ability. Either way, the majority of the pure ingroup effect can be attributed to discrimination that participants are aware of, while a fifth of the original size (0.010**) represents discrimination beyond perceived quality.³¹ Similar reductions in effect sizes are seen across hypotheses. The homophily-driven positive effect from ECB experts becomes more significant (5% level instead of 10%) but drops to roughly a third of its original size. Similarly, the increased signal use of NCB experts compared to ECB experts falls to slightly less than a third, while remaining significant at the 5% level. Therefore, while reported perceived quality explains much of the increased signal uptake, a remainder of ingroup effects persists.

Trust could be another driver of signal use and is in fact important. Columns (4) and (5) show that for ECB policymakers (homophily effect), accounting for trust does not fully explain differences in belief updating, though it explains more than half of the effect.³² In contrast, the increased signal use from the NCB expert can be entirely explained by trust; when interacting signal uptake coefficients with trust, there is no significant difference between the NCB and ECB experts. Thus, trust fully explains the differences in signal use between institutional experts.

³¹This is a conservative estimate, imposing that participants fully incorporate their own perceived messenger ability beyond the provided signal precision.

³²Note that the ingroup effect for ECB experts (H2) is insignificant when restricting the sample to only individuals where trust in representative policymakers is available. Hence, the comparison made here is based on the homophily effect (H3), where more observations are available.

1.3.6 Ingroup Policymakers Raise Information Reach but Not Attention

Beyond belief updating, the other dimension of messenger effects identified on Twitter is information availability. Does the increased volume of reporting translate to better reach? I find that it does: Participants in the experiment are better informed of ECB policymakers of their in- rather than their outgroup. Their probability of knowing these policymakers and receiving news about them increases by 27% and 29% respectively (See Appendix A.2.4 for details of the analysis). Thus, the ingroup is reached better.

Is the ingroup better informed because they pay more attention to information or simply by the higher availability? I causally evaluate paid attention within my experiment and find that attention to information, measured by clicking on “read more” buttons, remains unchanged across messenger treatments. Participants click on average on around 1.8 out of 3 buttons, independent of the messenger treatment (see Figure A.10). Therefore, the increased informedness on ingroup policymakers seems to be due to higher information availability, not increased attention to information.

1.3.6.1 Attention Is Endogenous

While attention to information does not vary between messengers, it instead responds to the inflationary environment, confirming that the experimental feature used to measure attention to information is successful. The randomization of messengers with inflation scenarios renders inflation scenarios’ inflation level and uncertainty exogenous. Therefore, causal effects of inflation and uncertainty on attention to information can be obtained. This is similar to estimating the extent of attention that is endogenous to the inflationary environment.

Higher inflation levels and greater inflation uncertainty significantly increase attention to information (see Table 1.3 or Figure A.10). A one percentage point increase in average historical inflation raises the likelihood of participants requesting any piece of information by 32.6 percentage points and raises the clicked buttons by almost 1 (0.921***).

Similarly, a one-unit increase in inflation uncertainty raises the probability of requesting any information by 24.3 percentage points and raises clicked buttons by 0.701***. These findings align with Cavallo et al. (2017), who find that individuals have weaker priors about inflation in low-inflation contexts, and with Weber et al. (2025), who conclude that attention to inflation is endogenous to its level. Besides confirming these results on the level of inflation, I further show that attention to information is endogenous to the uncertainty of inflation.³³

1.3.6.2 Attention Raises Ingroup Effects

While variation in the messenger does not impact attention to central bank communication, attentive participants—who request additional information—exhibit stronger ingroup effects across hypotheses (see A.13). Notably, these effects disappear when limiting the

³³Endogenous attention to information is rationalized by the costs of being mistaken about inflation increasing with the level of inflation in the real world. In my controlled setting, where participants are only incentivized to minimize forecasting errors, this is not the case: Expected forecasting errors should not depend on the level of inflation, only on the uncertainty.

Table 1.3. Endogenous Attention

Dep.:	OLS Attention (Continuous)	Probit Attention (Binary)	Marginal Effects Attention (Binary)
Inflation Level	0.921*** (0.090)	0.864*** (0.136)	0.326*** (0.050)
Inflation Uncertainty	0.701*** (0.077)	0.644*** (0.114)	0.243*** (0.042)
Signal Precision	0.179*** (0.033)	0.164*** (0.048)	0.062*** (0.018)
Cons	-1.117** (0.467)	-2.193*** (0.439)	
<i>Messenger Treatments</i>			
Outgroup Expert	0.059 (0.063)	0.047 (0.091)	
ECB Ingroup Expert	-0.053 (0.063)	-0.063 (0.091)	
ECB Outgroup Expert	-0.029 (0.063)	-0.042 (0.091)	
ECB Expert	-0.064 (0.063)	-0.050 (0.091)	
NCB Expert	-0.055 (0.063)	-0.040 (0.091)	
Treatment Order	✓	✓	✓
Individual-FE	✓		
R-squared	0.692		
Pseudo R-squared		0.023	
N	2,394	2,394	2,394

Notes: The dependent variable in column (1) is the number of additional information pieces requested (i.e., buttons clicked), in column (2) it is a binary variable, equaling 1 if at least some additional information piece is requested and 0 otherwise. Column (3) shows marginal effects of probit shown in column (2). Inflation level refers to the average inflation level of an inflation scenario, inflation uncertainty refers to the standard deviation of inflation of the 10 periods making up an inflation scenario. Reference category for messenger treatments is the Generic Ingroup Expert treatment. Stars correspond to the following p-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

sample to forecasting tasks where no buttons are revealed (i.e., “no attention”). Conversely, the effects intensify as individuals pay more attention (i.e., reveal more buttons). This underscores the importance of participant compliance with information treatments in experiments, as highlighted by Knotek et al. (2024). Details are in Appendix A.2.3.

1.3.7 Discussion of Experimental Findings and Policy Implications

The experimental findings demonstrate positive ingroup effects on information processing, which somewhat diminish within the ECB context. This might raise questions about how these results align with real-world evidence from Twitter and what they imply for policymaking.

Policymakers embody a multitude of characteristics beyond nationality, making real-world evidence difficult to square with the clean results from the experimental setting. A moderating factor for the pronounced results on belief updating seen on Twitter is certainly the increased availability of signals (i.e., news supply), as shown in Sections 1.2.3 and 1.3.6. Ingroup policymakers are significantly better known, but also significantly followed more

in the news than outgroup policymakers.

In sum, sharing nationalities with the communicating policymaker positively impacts the likelihood of receiving information and the extent to which novel information is used in belief updating. Overall, this makes central bank communication more effective for ingroup agents. So how can policymaking benefit from this? Reduced-form experimental evidence highlights that if central bank communication wants to target certain groups of the population, it can be beneficial to delegate communication to selected board members, due to their higher perceived technical abilities. Similarly, because of higher trust, communicating through the Eurosystem of National Central Banks improves the effectiveness of central bank communication. Even though attention to information does not causally increase for the ingroup, they are more likely to receive news about ingroup policymakers. To evaluate optimal communication policy, I develop a stylized coordination game that incorporates messenger effects on information availability and processing. This model provides normative recommendations for communication strategies in contexts with heterogeneous messengers and receivers.

1.4 Modeling Optimal Communication

Building on the empirical findings, this section develops a model with strategic complementarity to evaluate optimal public communication that maximizes social welfare. Extending Morris and Shin (2002), a “beauty contest” with private and public signals, the model incorporates the messenger’s role in public communication. Based on shared characteristics with the messenger, agents are classified as either in- or outgroup agents, which affects their availability and processing of the public signal. The social value of public communication is modeled as a “beauty contest”, reflecting the dual impact of public information disclosure: Agents respond to public signals due to both the information provided about a fundamental and their expectations of others’ reactions, creating strategic complementarity as they align their actions with both the fundamental and others’ choices (coordination motive).³⁴

Two communication policies are evaluated: a disclosure policy, affecting the precision of public signals, and a novel delegation policy, which involves choosing messengers to influence the composition of in- and outgroup agents, while keeping the single public signal constant.

Interestingly, I find that delegating communication to harness ingroup effects is not always socially desirable. Increasing the share of ingroup agents raises social welfare when public information is precise and coordination motives are moderate. By contrast, when private information is better and coordination is strong, the presence of outgroup agents is indeed desirable as their availability and processing of public information can mitigate potential welfare losses. Thus, adjusting the composition of ingroup and outgroup agents through delegation should be used strategically. Thus, delegation serves as a powerful additional policy tool to maximize welfare (conditional on messenger expertise), complementing disclosure policies.

The model is kept general and is applicable to any public communication with heteroge-

³⁴Examples of this coordination motive are firms setting prices considering their competitors’ prices, or agents negotiating wages similar to others.

neous messengers and receivers, including monetary policy as well as other policies, such as climate, health, fiscal or education. While nationality within the Euro area is chosen as an application to build intuition, the model extends to any dimension of heterogeneity, such as gender, ethnicity, socio-economic background, or economic expertise.

1.4.1 The Model Set-Up

A continuum of agents $i \in [0, 1]$ receive private and public signals and aim to align their actions with both an unknown fundamental and the actions of others (the coordination motive inspired by “beauty contest”). My key innovation is distinguishing between two *agent types* $h \in \{g, o\}$ —ingroup (g) and outgroup (o)—based on their alignment with the messenger’s characteristics ($\theta_i = \theta_m$ for ingroup, $\theta_i \neq \theta_m$ for outgroup). Reflecting my empirical findings, these types differ in the availability and processing of public information.

1.4.1.1 Information Structure

The fundamental $x \sim \mathcal{N}(\mu, \tau_x^{-1})$ represents an unobserved variable, such as inflation. Each agent $i \in [0, 1]$ predicts x based on private and public signals.

$$\text{Private signal: } y_i = x + \epsilon_{iy}, \quad \epsilon_{iy} \sim \mathcal{N}(0, \tau_y^{-1}) \quad (1.3)$$

$$\text{Public signal: } Y = x + \epsilon_Y = x + \epsilon_z + \epsilon_V, \quad \epsilon_z \sim \mathcal{N}(0, \tau_z^{-1}), \quad \epsilon_V \sim \mathcal{N}(0, \tau_V^{-1}) \quad (1.4)$$

The central bank’s own information about the fundamental ($z = x + \epsilon_z$) has a fixed noise component ϵ_z . As part of its *Disclosure Policy*, the central bank decides how much of this information to disclose by controlling the additional noise term ϵ_V . This affects the overall precision of the public signal τ_Y beyond τ_z . Full disclosure corresponds to $\tau_V \rightarrow \infty$, while complete opacity is $\tau_V \rightarrow 0$, and partial disclosure is any interim case, $\tau_V \in \mathbb{R}^+$. Accounting for both disclosure and noise in the public information, the observed public signal can be written as $Y \sim \mathcal{N}(x, \tau_Y^{-1})$, where $\tau_Y^{-1} = \tau_V^{-1} + \tau_z^{-1}$.³⁵

The second decision the central bank makes is on its *Delegation Policy*. By choosing the messenger(s) of communication, the central bank sets the share of ingroup agents in the economy, denoted α . Delegating communication assumes a single public signal Y and, unlike the disclosure policy, does not aim at changing the precision τ_Y but influences whether different agents receive and how they process the public signal.

Public Signal Availability: The public signal is not always available to all agent types—it is not strictly common knowledge. This deviates from coordination games in the literature, except for Cornand and Heinemann (2008) who restrict availability of public signals, but who do not let this depend on agent type and signal size.

Ingroup agents always receive the public signal and are therefore always informed. Outgroup agents, however, may or may not receive the public signal depending on whether its magnitude $|Y|$ exceeds their individual threshold d_o . Each outgroup agent $o \in [0, 1]$ has a threshold d_o drawn from a truncated normal distribution on the interval $[0, \infty)$, denoted $d_o \sim \mathcal{N}_+(0, 1)$ (see Appendix A.4.1 for details). An outgroup agent observes and acts on Y

³⁵Note that ϵ_V and ϵ_z are independent of each other, x , and of ϵ_{iy} .

only if $|Y| \geq d_o$; otherwise, the agent relies solely on their private signal. This captures the idea that agents obtain public signals indirectly, for example, through media outlets that report only sufficiently “newsworthy” signals (Nimark and Pitschner, 2019). Larger signals are more likely to be reported, increasing the likelihood that outgroup agents receive them.

The probability that an outgroup agent observes a public signal Y depends on its realized value ($\mathbb{P}(|Y| \geq d_o)$). Consequently, the fraction of informed outgroup agents is:

$$A = 2\Phi(|Y|) - 1, \quad (1.5)$$

where Φ is the standard normal cumulative distribution function. The fraction of outgroup agents who do not observe the signal is $1 - A$. As $|Y|$ becomes very large, A approaches 1, meaning nearly all outgroup agents observe large signals, whether positive or negative.

Belief Updating: All agents have a diffuse common prior that x is normally distributed with mean zero and zero precision. Agents follow Bayesian belief updating for private signals, which they receive regardless of the size of the signal. For ingroup agents, this is also the case for the public signal. However, informed outgroup agents incorporate *Information Resonance* in their belief updating (Malmendier and Veldkamp, 2022). The overall weight applied to a public signal is called the *Relevance Weight* ω_{im} and is defined as:

$$\omega_{im} = \rho_{im} \tau_Y = (2 - 2\Phi(\chi \|\theta_i, \theta_m\|)) \tau_Y, \quad (1.6)$$

where ρ_{im} is the *Resonance Weight*, which defines the resonance that a signal coming from m has for receiver i . $\|\theta_i, \theta_m\|$ is the Euclidean distance between the agent’s characteristics θ_i and the messenger’s characteristics θ_m , and χ captures the sensitivity to mismatching characteristics.³⁶ The normal cumulative distribution function Φ equals 0.5 at 0, and hence the expression $2 - 2\Phi(\cdot)$ takes on a value of 1 at zero and then declines quickly towards zero.

Scaling the resonance weight ρ_{im} with the (true) accuracy of the signal τ_Y reflects the idea that agents will not weight signals that have no information content at all.

The information sets are summarized as:

$$\Omega_i = \begin{cases} \{y_i, Y\} & \text{if } h = g, \\ \{y_i, Y\} & \text{if } h = o \text{ and } |Y| \geq d_o, \\ \{y_i\} & \text{if } h = o \text{ and } |Y| < d_o. \end{cases} \quad (1.7)$$

In addition to the signals, all agents know the share of ingroup agents α , and the fraction of informed outgroup agents A . Each informed agent believes that all agents—both ingroup and outgroup—process any received public signals in the same way they do.

Agents’ expected values of the fundamental x are:

³⁶If $\chi = 0$: agents learn from everyone without discounting anyone’s information $\rho_{im} = 1$. If χ is very large: agents disregard any information from messengers that do not exactly match their characteristics $\rho_{im} < 1$. $0 < \chi < 1$ represents cases in-between.

$$\text{Ingroup Agents: } E_{ig}[x | y_i, Y] = \frac{\tau_y y_i + \tau_Y Y}{\tau_y + \tau_Y}, \quad (1.8)$$

$$\text{Outgroup Agents: } E_{io}[x | y_i, Y] = \frac{\tau_y y_i + \rho_{im} \tau_Y Y}{\tau_y + \rho_{im} \tau_Y} \quad \text{if } |Y| \geq d_o, \quad (1.9)$$

$$E_{io}[x | y_i] = y_i \quad \text{if } |Y| < d_o. \quad (1.10)$$

1.4.1.2 Payoffs

Agents $i \in [0, 1]$ choose action $a_i \in \mathbb{R}$ to maximize their payoff $u_i \in \mathbb{R}$. Such an action could reflect decisions about consumption, savings or investments. Their payoff depends on their own action, strategic complementarities, and the state of the exogenous fundamental $x \in \mathbb{R}$:

$$u_i = -(1 - r)(a_i - x)^2 - r(a_i - \bar{a})^2, \quad (1.11)$$

where $\bar{a} = \int_0^1 a_i di$ is the average action. The term $-(a_i - x)^2$ represents the quadratic loss between the agent's action a_i and the exogenous fundamental state x . It reflects the agent's desire to align their action with the fundamental state. $-(a_i - \bar{a})^2$ captures the deviation of agent i 's action from the actions of all other agents a_j . This represents the "beauty contest" term: agents care about coordinating their actions with the average action of others.³⁷ r gives the weight on the coordinating, second-guessing motive.

1.4.1.3 Social Welfare

Social welfare is defined as the (normalized) average of individual utilities:

$$W(a, x) \equiv \frac{1}{1 - r} \int_0^1 u_i(a, x) di = - \int_0^1 (a_i - x)^2 di. \quad (1.12)$$

The social planner cares only about agents' distance to the fundamental x since the coordination motive introduces a social inefficiency, where agents' desire to align with others distorts individual actions without improving outcomes based on the true fundamental state.

1.4.1.4 Agents' Actions

While the social planner only cares about keeping all agents' actions close to the state x , the agent's action is determined by the information available to her and the first-order condition:

$$a_{ih} = (1 - r)E[x | \Omega_i] + rE[\bar{a} | \Omega_i], \quad (1.13)$$

³⁷The former term is akin to an agent's aim to forecast inflation correctly so as to avoid over- or under-consumption. The latter term of aligning forecasts with others can be viewed as to avoid disruptions or mismatches in wages, contracts, and prices, even if the forecast is not perfect. This coordination reduces uncertainty in economic interactions like wage-setting and price adjustments, ensuring smoother participation in the economy.

Neither the central bank nor the agents observe other agents' actions. Instead, agents form expectations about the average action, based on their information sets and requiring the determination of equilibrium signal extraction weights. To start with, the expectation of others' private signals is their expected state of the economy. For ingroup agents this is:

$$E_{ig}[y_j | y_i, Y] = E_{ig}[x | y_i, Y] = \frac{\tau_y y_i + \tau_Y Y}{\tau_y + \tau_Y}, \quad (1.14)$$

where $j \neq i$, hence y_j represents other agents' private signals. For informed outgroup agents this is:

$$E_{io}[y_j | y_i, Y] = E_{io}[x | y_i, Y] = \frac{\tau_y y_i + \rho_{im} \tau_Y Y}{\tau_y + \rho_{im} \tau_Y}. \quad (1.15)$$

For uninformed outgroup agents this is simply their private signal, which coincides with their expected value of the fundamental:

$$E_{io}[y_j | y_i] = E_{io}[x | y_i] = y_i. \quad (1.16)$$

1.4.1.5 Timeline

The game consists of two stages. First, the central bank communicates, having decided on their policy to disclose (τ_Y) and their policy to delegate (α) in order to maximize expected welfare. Second, agents receive their signals and choose their actions a_i to maximize expected utility based on their information sets.

1.4.2 Equilibrium

An equilibrium of the game consists of strategies for the central bank and the agents such that no player has an incentive to deviate.

1.4.2.1 Derivation of Equilibrium Actions

Suppose that all agents of type h follow a linear strategy:

$$a_{ih} = \kappa_h y_i + (1 - \kappa_h) Y. \quad (1.17)$$

Ingroup and Informed Outgroup Agents All informed agents of type h choose the same optimal weight κ_h in equilibrium. The optimal weight depends on an agent's expectations about the behavior of other agents. The shares of the population in the ingroup (α), outgroup ($1 - \alpha$), and informed outgroup (A) are common knowledge. However, agents assume all informed agents process information identically, unaware of any biases in others' or their own updating.³⁸

³⁸Assuming a lack of awareness of biases aligns with the inherent definition of bias and is consistent with its usage in the literature (e.g., Angeletos and Huo (2021); Gust and Lopez-Salido (2024)).

Lemma 1. *The optimal signal extraction weights κ_i for ingroup and informed outgroup agents are given by:*

$$\kappa_g^* = \frac{\tau_y q}{\tau_Y + \tau_y q} \quad \text{and} \quad \kappa_o^* = \frac{\tau_y q}{\rho_{im} \tau_Y + \tau_y q}, \quad (1.18)$$

where $q = (1 - r + r(1 - \alpha)(1 - A))$. *Proof in Appendix A.4.2.1.*

Equilibrium signal extraction weights κ_o^* and κ_g^* differ only by how the public signal's precision is being perceived (i.e., ingroup agents use τ_Y instead of $\rho_{im} \tau_Y$). The equilibrium actions of informed agents thus are:

$$a_{ig}(y_i, Y) = \frac{\tau_Y Y + \tau_y q y_i}{\tau_Y + \tau_y q} \quad \text{and} \quad a_{io}(y_i, Y) = \frac{\rho_{im} \tau_Y Y + \tau_y q y_i}{\rho_{im} \tau_Y + \tau_y q}. \quad (1.19)$$

Uninformed Outgroup Agents An uninformed outgroup agent can solely rely on their private signal:

$$a_{io}(y_i) = y_i. \quad (1.20)$$

Average Action The average action in equilibrium is:

$$\begin{aligned} \bar{a} &= \alpha a_{ig} + (1 - \alpha) A a_{io} + (1 - \alpha)(1 - A) a_{io}(y_i) \\ &= x \left[\alpha \kappa_g^* + (1 - \alpha) A \kappa_o^* + (1 - \alpha)(1 - A) \right] + Y \left[\alpha(1 - \kappa_g^*) + (1 - \alpha) A(1 - \kappa_o^*) \right]. \end{aligned} \quad (1.21)$$

Note that my model nests Morris and Shin (2002). This is the case in two special cases:

1. $\alpha = 1$: All agents are ingroup agents, leading to homogeneous agents who update without bias.
2. As $|Y| \rightarrow \infty$ and $\chi = 0$: All outgroup agents receive the signal ($A \rightarrow 1$) and update without bias ($\rho_{im} = 1$), eliminating differences between agent types.

1.4.2.2 Uniqueness of Equilibrium

Proposition 1. *The equilibrium strategies derived constitute the unique linear equilibrium of the game. Proof in Appendix A.4.2.2.*

1.4.3 Welfare Implications and Policy Analysis

Having derived the equilibrium, I now assess the welfare implications of the central bank's communication policies.

1.4.3.1 Social Welfare Computation

Expected social welfare, accounting for ingroup and outgroup agents, is:

$$\begin{aligned}
E[W(a, x)|x] &= -E \left[\int_0^1 (a_i - x)^2 di \right] \\
&= -\alpha \int_0^1 E \left[(a_{ig}(y_i, Y) - x)^2 \right] di - (1 - \alpha) A \int_0^1 E \left[(a_{io}(y_i, Y) - x)^2 \right] di \\
&\quad - (1 - \alpha)(1 - A) \int_0^1 E \left[(a_{io}(y_i) - x)^2 \right] di \\
&= -\alpha \left[\kappa_g^2 E(\epsilon_{iy}^2) + (1 - \kappa_g)^2 E(\epsilon_Y^2) \right] \\
&\quad - (1 - \alpha) A \left[\kappa_o^2 E(\epsilon_{iy}^2) + (1 - \kappa_o)^2 E(\epsilon_Y^2) \right] - (1 - \alpha)(1 - A) E(\epsilon_{iy}^2) \\
&= -\alpha \frac{\tau_Y + \tau_y q^2}{[\tau_Y + \tau_y q]^2} - (1 - \alpha) A \frac{\rho_{im}^2 \tau_Y + \tau_y q^2}{[\rho_{im} \tau_Y + \tau_y q]^2} - \frac{(1 - \alpha)(1 - A)}{\tau_y}.
\end{aligned} \tag{1.22}$$

How does the precision of information affect expected equilibrium social welfare? Conducting comparative statics on Equation (1.22) highlights the implications of raising the precision of public signals—akin to the disclosure policy—for welfare and policy design.³⁹

1.4.3.2 The Precision of Public Information: The Disclosure Policy

Recall that the central bank can fully disclose its information, thereby increasing the precision τ_Y of the public signal. The upper bound of Y 's precision, achieved under full disclosure ($\tau_Y \rightarrow \infty$), is τ_z , which represents the precision in the central bank's information itself. Conversely, under complete opacity, $\tau_Y = 0$ (since $\tau_Y \rightarrow 0$), meaning that Y is entirely noise and carries no meaningful information about the fundamental.

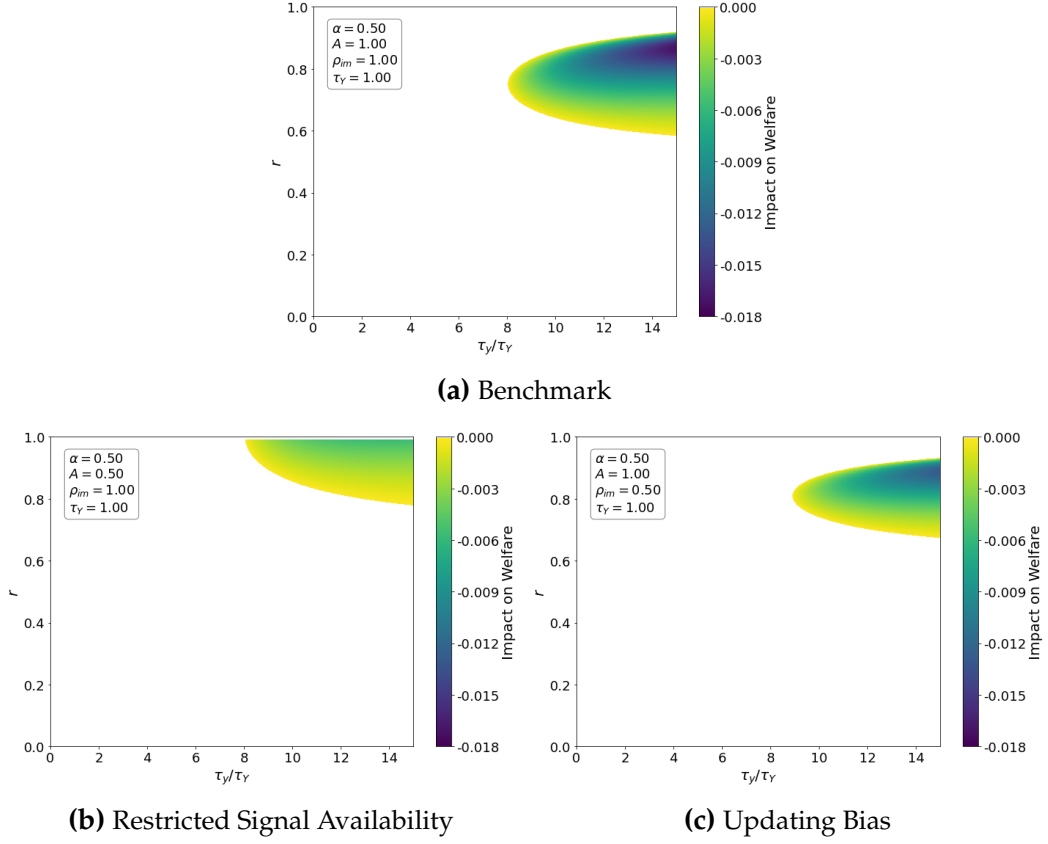
Proposition 2. *Increasing the precision of the public signal (τ_Y) improves social welfare only if the public signal is sufficiently precise relative to private signals and if the coordination motive r is not too high. Derivation in Appendix A.4.2.4.*

The canonical finding by Morris and Shin (2002) is that welfare is not necessarily increasing in the precision of the public signal—only if τ_Y is sufficiently large, especially compared to τ_y . This is because disclosing noisy public information causes agents with a desire to coordinate to neglect their private signals, which diminishes welfare when private information is better. Since my model nests Morris and Shin (2002), I use this as a benchmark in Figure 1.6a. Here, disclosure harms welfare when the coordination motive is sufficiently high ($r > 0.5$), and private signals are relatively precise compared to the public signal. The other two panels in Figure 1.6 highlight what happens to welfare if messenger effects are introduced one by one.

First, the outgroup's restricted signal availability reduces the parameter combinations for which disclosure harms welfare. Higher coordination is now required for disclosure

³⁹For completeness, the impact on welfare of raising the precision of private signals is discussed in Appendix A.4.2.3.

Figure 1.6. Disclosure's Effect on Social Welfare



Notes: The plots visualize Equation (A.19) by plotting the impact of increasing public signal precision (disclosure) on social welfare for different value combinations of parameters r and $\frac{\tau_y}{\tau_Y} = \tau_y$ since τ_Y is set to 1. White areas indicate strictly non-negative impact on welfare. Any other color indicates negative marginal effects on welfare. Panel (a) serves as the benchmark of all agents receiving the public signal and updating without a bias (equivalent to only ingroup agents as in Morris and Shin (2002)). Panel (b) shows the case where outgroup agents do not receive all signals but update without bias. Panel (c) shows the case where outgroup agents receive all signals but update with a bias.

to lower welfare (see Figure 1.6b). This is because not all public signals are universally observed. Instead, signal availability depends on signal size, where the outgroup is less likely to observe smaller signals (lower A). This can mitigate welfare loss from disclosure by limiting the over-reliance on relatively imprecise public information.

Second, the biased updating of outgroup agents further eliminates parameter combinations that render disclosure welfare-harming, as illustrated in Figure 1.6c. This occurs because outgroup agents downweight the public signal, which mitigates the over-reliance on imprecise public information in high coordination environments.

Thus, outgroup agents mitigate potential welfare losses from disclosure. Their presence acts as a buffer against scenarios where disclosure of public information could have detrimental welfare effects. Considering this, whenever the central bank has noisy information, it could delegate communication to optimally set the level of in- and outgroup agents (adjusting α),

in tandem with choosing its optimal disclosure policy (τ_V). This raises the question of how the central bank should optimally balance delegation and disclosure policies to maximize welfare.

Corollary 1. *The presence of outgroup agents can mitigate potential welfare losses from increased disclosure when public information is noisy.*

1.4.3.3 The Composition of In- and Outgroup Agents: The Delegation Policy

The central bank may decide to delegate communication of the single public signal from the default messenger (typically the president) to one or more messengers of varying characteristics. Delegation controls the share of ingroup agents (α); either by selecting multiple messengers communicating the same signal, or by selecting another messenger whose characteristics resonate with a larger share of the economy. Reduced-form evidence of this paper suggests the positive ingroup effect on belief updating carries over to two concrete examples of delegation policy: through other board members with varying characteristics, or through governors of national institutions.

For the central bank to decide on the policy tool of delegating communication, it needs to determine the welfare-maximizing share of ingroup agents, $0 \leq \alpha^* \leq 1$. This optimal share of ingroup agents depends on the coordination motive r , the availability of the public signal to the outgroup A (thus, the size of the signal $|Y|$), the size of outgroup agents' updating bias ρ_{im} and the precisions of signals τ_y and τ_V .

Proposition 3. *An optimal share of ingroup agents (α^*) exists and depends on r , signal precision, signal size, and bias in information processing. Derivation in Appendix A.4.2.5.*

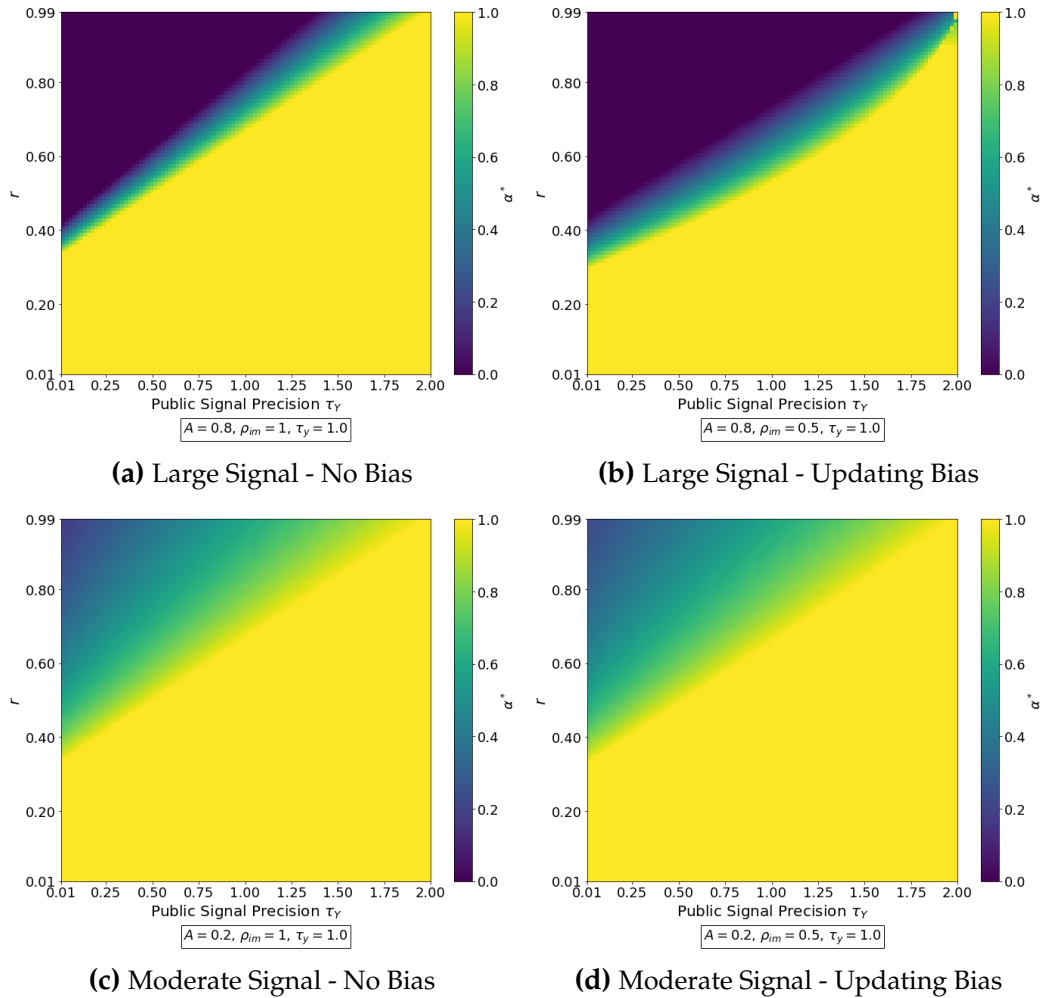
An explicit analytical solution for α^* is not feasible. Instead, I use numerical solutions to illustrate how α^* varies with parameters and discuss special cases.

For the special case, in which the outgroup faces restricted signal availability but updates without bias, $\alpha^* = \min \left\{ 1, \frac{\tau_V + \tau_y}{3r\tau_y} \right\}$. Meaning, the optimal share of ingroup agents α^* is less than one so long as the private signal is sufficiently precise compared to the public signal and there is sufficiently strong strategic complementarity (formally, $\frac{\tau_V}{\tau_y} = 3r - 1$). $r < \frac{1}{3}$ always results in the corner solution of all agents optimally being ingroup agents ($\alpha^* = 1$). This result is in line with the optimal publicity result of Cornand and Heinemann (2008). Intuitively, over-reaction to a relatively noisy public signal is alleviated by outgroup agents not receiving smaller signals. The presence of both ingroup and outgroup agents creates a dynamic where the negative impact of ingroup agents' overreaction is outweighed by the positive effect of enhanced coordination on x .

Introducing an updating bias complicates the expression for α^* . Figure 1.7 shows α^* numerically, illustrating optimal delegation across different levels of coordination and public signal precision. The left column (Panels 1.7a and 1.7c) represents cases of outgroup agents updating without bias but not receiving all signals. The right column (Panels 1.7b and 1.7d) adds biased updating ($\rho_{im} = 0.5$). Top and bottom rows compare large signals (received by 80% of outgroup agents) to moderate signals (received by only 20%). Regardless of signal size, the more precise public information is relative to private information, the

higher the optimal share of ingroup agents. Intuitively, this is because if public information is very precise, it is beneficial if agents receive this information, and fully incorporate it in their expectations. By contrast, when public information is relatively noisy, it is beneficial to have fewer ingroup agents, so α^* drops. This is especially pronounced for large signals, when the outgroup's bias to belief updating plays a substantial role, lowering the optimal share of ingroup agents for more parameter combinations than in the absence of bias. Meaning, given the bias, fewer ingroup agents may be socially desirable, even in cases of weaker strategic complementarities and noisier public signals than if outgroup agents updated without bias. When signals are moderate, the optimal share of ingroup agents drops to a lesser extent. In this case, the belief updating bias makes little difference to the optimal share of ingroup agents.

Figure 1.7. Optimal Share of Ingroup Agents



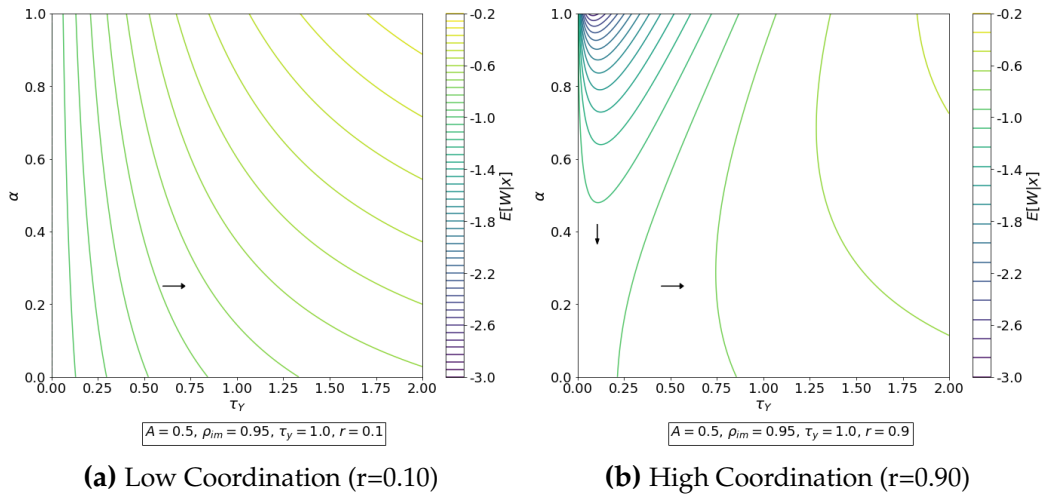
Notes: Plots show the welfare-maximizing level of ingroup agents α^* for different combinations of coordination (r) and public signal precision (τ_Y). The top row plots show α^* for large signals (corresponding to $A = 0.8$), and bottom row plots show the same for moderate signals ($A = 0.2$). Severe updating bias corresponds to $\rho_{im} = 0.5$. Since $\tau_y = 1$, the public signal is more precise than the private signal whenever $\tau_Y > 1$ and noisier for $\tau_Y < 1$.

A Potential Caveat of Delegating Communication: Loss in Expertise Delegating communication away from a single president, who is presumably be the best-suited candidate to communicate, might come at the cost of public signal precision. Such precision losses arise if the messenger conveys less expertise than the original messenger. This potential trade-off must be accounted for when deciding on whether to delegate communication. Appendix A.4.3 discusses this formally.

1.4.4 Policy Implications: When to Delegate?

In economies with heterogeneous agents and messengers, both disclosure and delegation policies contribute to social welfare. How should a central bank choose between these policies, or a mix of them? I evaluate disclosure (raising τ_Y) against delegation (raising α) and illustrate the social welfare contours for environments of high and low coordination (Figure 1.8). The share of informed outgroup agents is set to 50%, and their updating bias is set to $\rho_{im} = 0.95$, reflecting the pure ingroup effect this paper finds empirically. To showcase the role of coordination, r is set to take on rather extreme values, which are $r = 0.90$ and 0.10 , respectively.

Figure 1.8. Social Welfare Contours



Notes: Plots show social welfare contours in (τ_Y, α) -space. τ_y is set to 1. Arrows indicate the direction in which welfare is increasing. Panel (a) shows a case of insufficient strategic complementarity ($r = 0.10$). Panel (b) shows a case of sufficient strategic complementarity ($r = 0.90$).

1. When coordination is low and public signals are precise: Full disclosure and maximizing the share of ingroup agents ($\alpha^* = 1$) is optimal.

When public information is relatively noisier than private information ($\tau_y > \tau_Y$), delegating communication is always optimal ($\alpha^* = 1$), and so is full disclosure ($\tau_Y^* \rightarrow \infty$). Therefore, the central bank should delegate communication to maximize the share of ingroup agents, and disclose all its information (see Figure 1.8a). When public information is noisier than private signals, welfare improvements of raising α are rather small. Larger increases in the share of ingroup agents are required for notable welfare improvements.

When private information is noisier ($\tau_y < \tau_Y$), delegating communication is optimal, as the corner solution of $\alpha^* = 1$ still applies. When the public signal is very precise, even small increases in the share of ingroup agents improve welfare substantially.

2. *When coordination is high and public signals are noisy: Limiting the share of ingroup agents can be beneficial to prevent over-reliance on imprecise public information.*

When public information is relatively noisier than private information ($\tau_y > \tau_Y$), the relative precision of the public signal must be sufficiently high for welfare to increase with the share of ingroup agents (see Figure 1.8b). If this is not achieved, delegating communication may be harmful to social welfare. With sufficiently high strategic complementarity it may be optimal to limit ingroup agents when public signals are noisy.⁴⁰ Such low levels of public signal precision may occur when the central bank's information is noisy (compared to private information), or when it decides not to disclose all or parts of its information. Note that when social welfare contours are downward-sloping, both delegation and disclosure harm welfare (see Figure 1.8b). The latter reflects the canonical finding of Morris and Shin (2002), while the former now identifies an additional policy tool for these circumstances: Alongside limiting disclosure, welfare can be raised by limiting delegation (e.g., through centralized communication or selecting homogeneous messengers), as this lowers the share of ingroup agents α .

Furthermore, there are circumstances in which disclosure increases social welfare, whereas delegation does not. This is the case when welfare contours are upward-sloping, until the public signal becomes relatively more precise. The reason for this is the heterogeneity in how agents process public signals: In these circumstances, outgroup agents assign greater weight to the more precise private information, as they perceive it to be relatively even better, and expect all other agents to do the same. This counteracts the coordination on relatively poorer public signals.⁴¹ As the quality of the public signal improves, close to when it becomes better than the private signal, both increasing and decreasing ingroup agents can be optimal. Thus, delegation is a *distinct* policy tool, different from traditional disclosure policy. Still, delegation should be decided on in conjunction with disclosure considerations.

When private information is noisier ($\tau_y < \tau_Y$), delegating communication proves beneficial as ingroup agents use the precise public signal more effectively. Consequently, to align actions and maximize overall social welfare, raising the share of ingroup agents (α^*) as much as possible is optimal.

When coordination is sufficiently low, delegating communication is beneficial provided that any precision losses from lower expertise are sufficiently small. Precision losses are less concerning when private signals are relatively noisy but can be detrimental when public

⁴⁰The necessary level of strategic complementarity for this depends on the updating bias and the availability of the signal to outgroup agents. It is $r > \frac{1}{3}$ when outgroup agents update without a bias (irrespective of signal availability), but lower than that when there is biased updating and signals are available to outgroup agents. For example, when signals are large ($A \rightarrow 1$) and there is a slight updating bias of $\rho_{im} = 0.95$, $r > 0.156$ is sufficiently high strategic complementarity. Larger signals generally return lower sufficient r , while stronger updating biases raise sufficient r .

⁴¹In addition, having many homogeneous outgroup agents—who interpret public signals identically—is optimal because of the coordination motive.

information is relatively noisier ($\tau_Y < \tau_y$). When coordination is high, communication should only be delegated if the public signal is ensured to be sufficiently precise. In such cases, even with full disclosure, perceived losses in expertise of a delegated messenger may harm welfare due to the strong coordination motive causing over-reliance on the noisy public signal.

In sum, the central bank's communication strategies on both disclosure and delegation affect social welfare through the social value of public information. Although increased disclosure of public information can enhance coordination among agents, it is not always welfare-improving, especially when the public signal is noisy. Outgroup agents buffer such potential welfare losses, as they rely less on public signals. Delegation—choosing the right messenger to set the composition of in- and outgroup agents—can raise welfare, but must be carefully balanced against possible reductions in expertise and thus perceived public signal quality. By jointly setting delegation and disclosure policies, while carefully considering the size and relative precision of public signals, central banks can optimize their communication strategies to maximize social welfare.

1.5 Conclusion

The messenger matters. To use communication as an effective policy tool, it is critical to consider the alignment between messenger and receiver characteristics. First, this paper empirically shows that the reach and impact of public policy communication are improved when the characteristics of the messenger and the audience match. Second, it uses theoretical methods to identify delegation—the strategic selection of the messenger—as an additional policy tool for optimal public communication.

The experiment provides causal support for two concrete ways of implementing a delegation policy in the context of central banks: delegating communication to regional central banks, or delegating communication to other board members with different characteristics.

The causally estimated messenger effects are driven by homophily rather than heterophobia, meaning that ingroup messengers are relied upon more, without outgroup messengers being disproportionately disregarded. The effects are largely explained by perceived quality and trust, and while messengers influence information availability, they do not affect attention to information, which instead responds to the inflationary environment, including uncertainty.

With a stylized coordination model, I then normatively assess optimal communication policy in light of the empirically estimated messenger effects. I account for the restricted public signal availability and biased information processing, and assess a novel delegation policy alongside the well-studied disclosure policy. While increasing the share of ingroup agents through delegation improves welfare when public information is precise, the presence of outgroup agents serves as a buffer against potential welfare losses from disclosing noisy public information, beyond restricting disclosure. Thus, a delegation policy complements disclosure in maximizing the social value of public information.

The messenger can serve as an additional, powerful policy tool. Strategically selecting who communicates public policy allows the maximization of social welfare, and therefore for the optimization of communication. This is true for policy communication with diverse

public audiences beyond monetary policy, including climate and fiscal policy. Similarly, beyond nationality, there are many dimensions of diversity that can be considered, such as age, gender, ethnicity, socio-economic background, or expertise.

Central Bank Communication With Non-Experts—A Road to Nowhere?^{*}

Abstract Central banks have intensified their communication with non-experts—an endeavour which some argue will fail. This paper studies English and German tweets about the ECB to show that its communication is received by non-experts, i.e. is not a road to nowhere. Following ECB communications, tweets often primarily relay information, become more factual and the views expressed more moderate and homogeneous. Some communications, such as Mario Draghi’s “Whatever it takes”, trigger a divergence in views. Also, tweets with negative, stronger or more subjective views are more likely retweeted, liked or replied to. Thus, Twitter also constitutes a platform for controversial discussions.

Keywords: Monetary policy, Central bank communication, Social media, Non-experts
JEL: E52, E58

2.1 Introduction

Central banks have travelled a long journey with their communication practices (Issing, 2019), towards ever-increasing transparency and intensified communication with experts, and especially with financial markets. The latest development in that journey is an increasing effort to communicate with non-experts, which has gained in importance following the global financial crisis, the subsequent use of unconventional monetary policy tools and the broadening of central bank mandates. New mandates and new tools require more

^{*}This chapter is co-authored with Michael Ehrmann and published in the *Journal of Monetary Economics* 127 (2022) 69–85 (Ehrmann and Wabitsch, 2022). Note that minor editorial revisions have been made to the text. This paper presents the authors’ personal opinions and does not necessarily reflect the views of the ECB or the Eurosystem. The authors would like to thank Philipp Gnan, Oleksiy Kryvtsov, Kilian Rieder, Peter Tillmann and anonymous referees for this journal and the ECB Working Paper Series as well as participants at an ECB seminar, the workshop “Central bank communication as a public good” and the Bank of Finland/CEPR conference “Monetary Policy Tools and their Impact on the Macroeconomy” for helpful comments.

explanation (also to the expert audiences); furthermore, these changes made monetary policy the focus of an intensifying, highly controversial public debate (Blinder et al., 2017). In addition, central banks saw an erosion of citizens' trust in them and their policies (Bergbauer et al., 2020). More communication with non-experts was thus needed; indeed, in her confirmatory parliamentary hearing, incoming European Central Bank (ECB) President Lagarde mentioned communication with the general public as one of the priorities of her presidency (Lagarde, 2019).

Reaching out to non-experts raises a number of challenges—up to the point that Blinder (2018) predicts that “central banks will keep trying to communicate with the general public, as they should. But for the most part, they will fail”. In contrast to experts, non-experts know less about central banks, might not be within reach, and will not necessarily respond as fast and visibly to central bank messages, making it more difficult to assess whether and how a certain message was received. In light of this, Haldane et al. (2020) call for “explanation, engagement and education”, or what they call the “3 E’s of central bank communication with the public”.

A necessary condition to explain and educate is that the central bank messages actually reach the non-experts. The literature on this central question is surprisingly scant. Ter Ellen et al. (2022) show that central bank communication requires news media as an intermediary, but can affect consumer confidence. This is in line with the findings of Lewis et al. (2019) that monetary policy surprises have instantaneous effects on consumer confidence in the United States. In contrast, Lamla and Vinogradov (2019), Lamla and Vinogradov (2022) and De Fiore et al. (2022) cast doubt on the ability of central banks to get through to consumers: while relatively more respondents in their surveys of US and UK consumers report to have heard news about the central bank following policy announcements (and more so if they are using Twitter), people’s expectations, in particular about inflation, are effectively unchanged.

This paper follows a novel avenue to observe the reaction of non-experts to central bank communication: it studies how they talk about the ECB in social media, by analysing tweets posted on Twitter. This approach has several advantages: it is based on real-life data (in contrast to lab or survey experiments, which impose that the central bank signals are received) which are available at high frequency (allowing causal analysis, in line with the assumptions underlying event studies) and on a continuous basis (therefore covering many communication events). Furthermore, it represents many individuals (more than listening events, lab or survey experiments), and it allows comparing non-experts and experts, as both are active on Twitter.

This paper studies tweets about the ECB for the years 2012 to 2018 in two languages, English (a *lingua franca*, the language spoken in financial markets and the language of most ECB communication) and German (the most widely spoken language in the euro area, which also hosted a particularly heated public debate about the ECB’s monetary policy).

The key findings are as follows. First, by assuming that non-experts write tweets about a variety of issues, and only occasionally tweet about the ECB or its policies, it is possible to differentiate experts from non-experts. Non-experts, who are considerably more likely to tweet during weekends, express stronger opinions, are more subjective in their views, and represent a much larger variety of views than the experts in the sample.

Second, the analysis identifies a dual role for Twitter—a channel for information transmission on the one hand and a platform for controversial and subjective discussions on the other hand. In general, ECB-related Twitter traffic is responsive to ECB communication. For most events, this effect is contained to the same day, with an increased number of Twitter accounts expressing less subjective views, in softer language and in a relatively more homogeneous fashion. This pattern suggests that tweets on these days mainly relay information about the ECB. In contrast, other ECB communications trigger a more persistent response and see a divergence of views—more pronounced so among tweets in German, and most prominently for former President Draghi’s “Whatever it takes” statement. These patterns suggest that Twitter also serves to controversially discuss the ECB’s policies.

Third, non-experts are less responsive to ECB communication than experts, in particular for events where tweets mainly relay information. For these events, tweets by non-experts tend to become more factual—their subjectivity becomes less pronounced and less dispersed, and there is a tendency towards more moderate views being expressed. In contrast, “Whatever it takes”, the prime example of an occasion where Twitter served as a platform for controversial discussions, yielded very similar reactions of experts and non-experts.

Fourth, Twitter’s role as a platform for controversial discussions also becomes apparent when analysing replies, retweets and likes: the likelihood to get a reply, retweeted or liked is higher for tweets that formulate their opinion in relatively strong and more subjective language, or that convey a negative view.

These findings have important implications for central banks. First, they suggest that central banks’ efforts to monitor the related social media traffic should be relatively granular and try to differentiate between expert and non-expert users, and furthermore between Twitter activity that serves primarily the purpose of information transmission and the more controversial discussions on Twitter. Second, monitoring the latter is particularly important, because the retweet and like analysis suggests that strong views and more subjective contributions are reposted more often, and hence are more influential in the discussion. At the same time, the results go against the view that central bank communication with non-experts is bound to fail because it does not reach the intended recipients. The ECB manages to reach out to non-experts. And central bank communication has the potential to make discussions in social media somewhat more factual and moderate. However, whether it ultimately succeeds in fostering trust, making central banks accountable, or influencing agents’ inflation expectations and behaviour remains an open issue.

2.2 Related Literature

This paper relates to two, so far largely unconnected, strands of literature. The first deals with social media in financial market and central bank-related contexts.¹ Korhonen and Newby (2019) and Conti-Brown and Feinstein (2020) document that central banks have

¹Twitter activity is studied in many other fields, too; reviewing this literature is beyond the scope of this paper. Still, it is worth highlighting studies of information diffusion in social media, as this has a bearing on the application in the current paper. For instance, Gorodnichenko et al. (2021) report that diffusion of information related to the 2016 Brexit referendum and the 2016 U.S. presidential elections is largely complete within one to two hours and shows signs of an “echo chamber”, with stronger interactions across agents with similar beliefs.

become very active on Twitter. Studies on market perceptions or market reactions using Twitter find that tweets contain relevant information—to predict stock market returns around FOMC meetings (Azar and Lo, 2016), to identify monetary policy shocks (Masciandaro et al., 2021), or to understand beliefs about monetary policy (Meinusch and Tillmann, 2017; Stiefel and Vivès, 2022; Lüdering and Tillmann, 2020). Tweets by U.S. president Trump about the Federal Reserve and its monetary policy have been found to have affected long-term inflation expectations and confidence of consumers (Binder, 2021) and to have led to a reduction in interest rates, suggesting that market participants priced in future rate cuts in response to Trump’s statements (Bianchi et al., 2023; Camous and Matveev, 2021; Tillmann, 2020).

This literature has provided ample evidence that the Twitter activity of central banks, financial market participants (or experts for that matter) and politicians contains useful information to study various aspects related to central banking. What is missing, to the best of our knowledge, is an analysis of Twitter activity by non-experts—the topic of this paper.

The second strand of literature to which this paper contributes is the recent research on central bank communication with non-experts. Prior to the global financial crisis, there was virtually no such research (Blinder et al., 2008). This is changing rapidly, often with the active involvement of central banks. For instance, the Bank of England augmented its Inflation Report with new layers of content aimed explicitly at speaking to a less-specialist audience and conducted controlled experiments to assess the impact of this change (Haldane and McMahon, 2018). In order to understand the determinants of trust, the ECB has been experimenting with changing the order of questions in its knowledge and attitudes survey among the general public (Angino and Secola, 2022), and the Bank of Canada has embarked on lab experiments to test the causal effects of central bank communication on economic expectations and their underlying mechanisms (Kryvtsov and Petersen, 2021). The literature has also been expanding rapidly due to the increasing popularity of randomised control trials (RCTs) which have been used by, for instance, Coibion et al. (2022) who study how different forms of communication influence inflation expectations, and D’Acunto et al. (2022) who investigate whether diversity in the committee helps reaching out to different population groups.

A clear message that emerges from these studies is that simple and relatable messages are more powerful in affecting beliefs or behaviours of non-experts (Bholat et al., 2019; Coibion et al., 2022; Kryvtsov and Petersen, 2021). This evidence is consistent with models in which agents have constrained capacity to collect and process information (Coibion et al., 2022).

This is an important message for central banks—after all, their communications are usually far from being a simple read: for instance, it requires around 13-15 years of formal education to understand the monetary policy statements of the ECB (Coenen et al., 2017).

Lab experiments and RCTs in surveys guarantee that the intended recipient receives the central bank signal—the participants get confronted with a message (or deliberately do not receive this message, to generate a control group), and then can react to it (or not). This is an advantage of these approaches, as it allows for a causal interpretation. At the same time, this is arguably also their largest downside—in real life, no one can guarantee that non-experts receive the central bank signal. The related evidence is sobering—non-experts tend to have little knowledge about central banks, and show little interest in keeping up to date with monetary policy issues (Van der Cruysen et al., 2015). Their inflation expectations

are not informed by the central banks' inflation targets (Kumar et al., 2015; Binder, 2017), depend on the information channel they use (Conrad et al., 2022) and are not responsive to monetary policy announcements in low-inflation environments (Coibion et al., 2023).

All of this implies that central bank communication with non-experts can substantially improve their knowledge and possibly also affect their expectations and behaviour—if the signals get through to them. This is where the current paper comes in—it studies to what extent the ECB manages to reach out to non-experts and whether and how the tone of the corresponding social media discussion can be affected by the ECB's communication efforts.

2.3 Data

This section describes the data used for the empirical analysis.

2.3.1 Tweets

The sample of tweets covers the time period from 2012, when the usage of Twitter in Europe has broadly stabilised,² to 2018, to ensure that the analysis is not affected by the changeover of the ECB presidency from Mario Draghi to Christine Lagarde in 2019. While a changeover in the leadership could generally make a difference, the fact that—unlike Mario Draghi—Christine Lagarde has a Twitter account and uses it actively likely implies a structural break in the data.

The tweets are filtered and scraped via Twitter's Advanced Search using the Python libraries "GetOldTweets" (Henrique, 2016) and "Twint" (Poldi and Zacharias, 2018).³ Tweets in English—as identified by Twitter's language filter—are collected if they contain "ecb", "european central bank" or "draghi" in the text, hashtag or username. For the sample of tweets in German, the language filter is set to German and tweets are collected if they contain "ecb", "ezb", "europäische zentralbank" or "draghi" in the text, hashtag or username. All searches are insensitive to capitalisation and special letters such as the umlaut. This results in over 4 million English tweets and almost 110,000 German tweets. The dataset covers what Twitter identifies as original tweets.⁴ In addition, it identifies the number of times each tweet receives a reply, gets liked and retweeted (but not whether the retweets themselves were retweeted any further).

The data are then cleaned in several steps to ensure that the final samples are indeed related to the European Central Bank, and written in English or German. The sample is limited to tweets from Twitter accounts that have tweeted at least 100 times. Also, to avoid circularity of the analysis, tweets that originate from the ECB's own Twitter account are dropped. Details of this cleaning procedure can be found in the Appendix.

²See [here](#).

³The data collection is not done in real time, but ex post. This implies that it retrieves all tweets that were publicly available online at the time of data collection, but does not discover tweets that got deleted in the meantime. This method of collecting tweets for scientific analysis has been used, inter alia, by Lan et al. (2019), by Tavazoe et al. (2017) or by Song and Miled (2017).

⁴This means that it covers new tweets and "improper" retweets (where content is copied from another tweet, but often a comment or remark is added) as well as the new content of quote retweets and replies (provided this new content meets the sample selection criteria), but not retweets.

The final dataset has more than 3.2 million original tweets in English, which generated more than 1.7 million retweets, not counting quote retweets⁵ (see Table B.1 in the Appendix). The sample of tweets in German is considerably smaller; there are only around 97,000 original tweets, which were retweeted around 32,000 times.

The top panel of Figure 2.1 tracks the evolution of tweets over time, and shows, first, that Twitter activity across the English and German subsample is highly correlated and, second, that Twitter activity peaks around major ECB decisions. The first peak corresponds to former President Draghi's "Whatever it takes" statement in July 2012. Also 2014 and 2015 show an elevated level of Twitter activity, following the comprehensive monetary policy easing starting in June 2014.⁶ This provides a first indication that ECB communications are an important determinant of ECB-related Twitter activity.

To put these data into perspective, the middle panel of Figure 2.1 plots the searches for ECB-related terms on Google, and the lower panel of Figure 2.1 reports the number of ECB-related articles in English-speaking newspapers. The three sources yield similar trends, suggesting that the collected tweets reflect well the general interest in ECB-related matters.

2.3.2 Content of Tweets

Besides the volume of tweets and retweets, the content of the tweets is of interest. To analyse the sentiment of the tweets, a dictionary approach is used.⁷ The Python library TextBlob is used for the English sample (Loria et al., 2014), its German extension for the German sample (Killer, 2015). Hence, the English lexicon is based on Princeton University's WordNet and the German lexicon on the German equivalent GermaNet.⁸ Three types of sentiment are calculated for each tweet: favourableness (i.e. tone of tweet), absolute favourableness (or sentiment strength) and subjectivity. Appendix Table B.3 lists example adjectives that return very high or low values for favourability and subjectivity.

Favourableness ranges from -1 to 1, where a higher value reflects a more positive sentiment. For instance, the words "awful" or "dreadful" are given a favourableness value of -1, "exceptional" or "marvelous" yield a value of +1. The absolute value of favourableness ranges from 0 to 1, where values closer to 1 reflect stronger sentiment. "Awful", "dreadful", "exceptional" or "marvelous" express strong views, with the absolute value of favourableness being +1. In contrast, neutral words such as "consistent" or "basic" carry low sentiment strength (absolute favourableness is 0). Subjectivity also ranges from 0 to 1, where higher values indicate less factual (more subjective) statements. "Nasty" or "terrible" yield high subjectivity values of 1, whereas the words "actual" or "contemporary" are given the lowest subjectivity value of 0. The coding takes into account if multiple words carrying sentiment

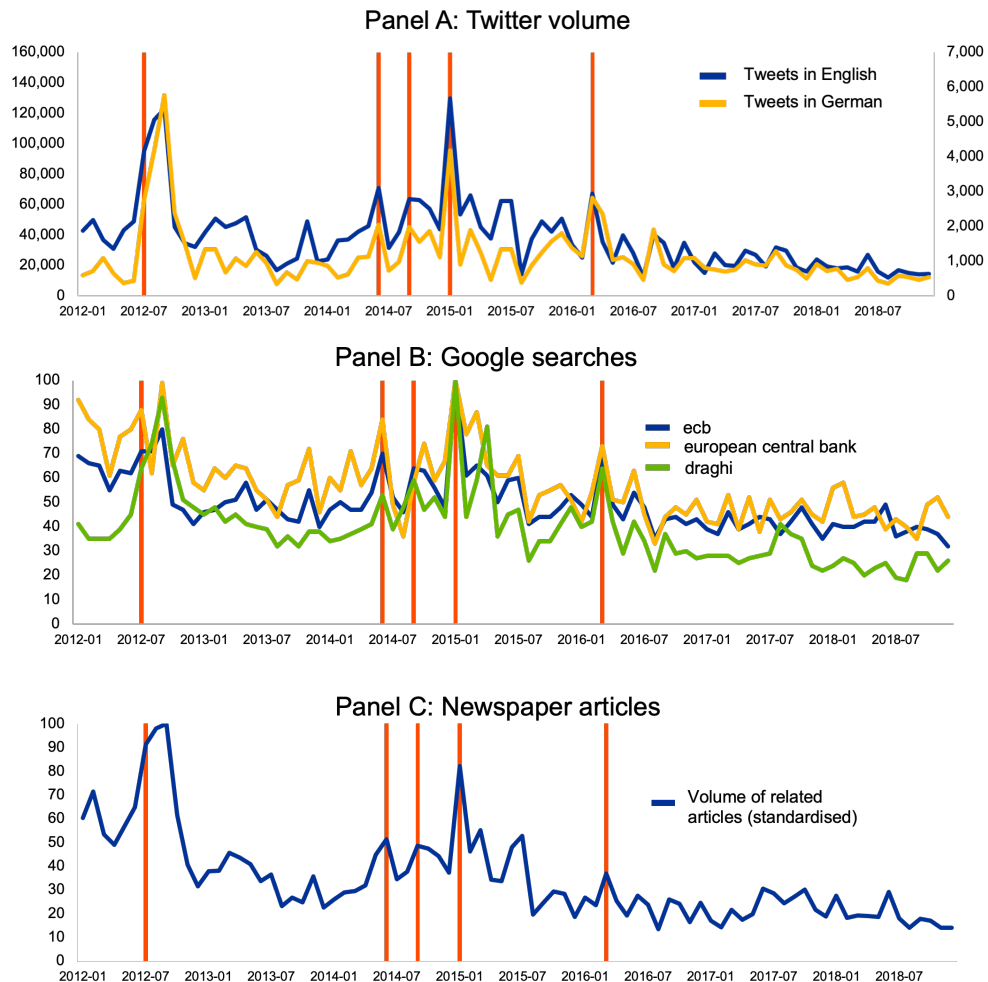
⁵Quote retweets are not counted, as they have only been introduced in 2015.

⁶For a detailed review, see Hartmann and Smets (2018).

⁷Shortcomings of this dictionary approach could be missing or misinterpreted words in the lexicon (e.g. "negative interest rate" returns a negative favourability value and a positive subjectivity score), unidentified words due to spelling mistakes, or sarcasm, irony or cynicism in the text. The latter should affect in particular the measurement of favourability, but less so absolute favourability and subjectivity. Accordingly, results for absolute favourability and subjectivity should be more robust and reliable.

⁸<https://wordnet.princeton.edu/> and <http://www.sfs.uni-tuebingen.de/GermaNet/>, respectively. The measures of sentiment cannot be directly compared across languages, in particular because the libraries do not use the same (translated) dictionaries.

Figure 2.1. Interest in the ECB: Number of Tweets, Google Searches, Newspaper Articles



Notes: Panel A: Monthly number of ECB-related tweets in English (left axis) and German (right axis). The vertical lines illustrate the timing of various ECB actions, namely July 2012: “Whatever it takes”; June 2014: Introduction of negative interest rates and credit-easing measures via targeted long-term refinancing operations, then complemented by and an asset-backed securities purchase programme; September 2014: Introduction of third covered bond purchase programme; January 2015: Expansion of asset purchase programme (APP), starting the public sector purchase programme (PSPP); March 2016: ECB lowers rates further and expands its APP considerably. Panel B: Monthly Google search popularity for the three search terms “ecb”, “european central bank” and “draghi”. Numbers represent search interest relative to the highest point on the chart for worldwide searches between 2012 and 2018. A value of 100 is the peak popularity for each term. Source: Google Trends (<https://www.google.com/trends>). Panel C: Number of newspaper articles related to the key terms “ecb”, “european central bank” and “draghi” in English. The sample is based on 3,075 different news outlets and on over 800,000 articles. Standardised values, where 100 is the peak of article volume. Source: Factiva DNA.

co-occur in one tweet (in that case, their average value of favourability and subjectivity is returned) and considers alterations of words (for instance, “good”, “not good” and “very good” yield favourability values of 0.7, -0.35 and 0.9, and subjectivity scores of 0.6, 0.6 and 0.8). Appendix Table B.2 reports summary statistics.⁹

2.3.3 Associated Twitter Accounts

For each user who is associated with at least one tweet about the ECB in the sample, the following information on the account is obtained using common web-scraping techniques: the date of account creation, the number of followers, and the number of overall tweets (“statuses”) that have been issued by the specific account since its creation.

ECB-related tweets in English originate from 255,336 accounts; tweets in German from 14,780 users.¹⁰ Most of the traffic stems from relatively few accounts: the top 5% of accounts generate 74% (61%) of tweets in English (German). Furthermore, the top 5% Twitter accounts are responsible for 93% (82%) of “liked” tweets, for 94% (85%) of retweeted tweets, and for 88% (82%) of replies to tweets in the English (German) sample. This suggests that there is a small number of Twitter accounts that account for most of the traffic, and even fewer constitute the most influential opinion-makers. Given the extremely unequal distribution, it is fair to use aggregate Twitter activity as representative for the expert population, be it media, financial market participants or economists and finance professionals. Previous literature has done exactly this, also finding that Twitter activity correlates well with financial market developments. At the same time, as argued below, it allows for a more differentiated analysis by isolating experts from non-experts.

2.3.4 ECB Communication Events

The following communication events by the ECB (sourced from the ECB’s website) are captured: announcements of monetary policy decisions and along with the accompanying press conference (monthly until 2014, eight times a year since 2015; 68 observations in the sample); the Economic Bulletin, which provides an overview of the economic and monetary information that forms the basis for the Governing Council’s policy decisions (released two weeks after each monetary policy meeting; 68 observations in the sample); the accounts of the monetary policy meetings (published since 2015, usually 4 weeks after the monetary policy meetings; 31 observations in the sample); tweets originating from the ECB’s institutional Twitter account on days without other ECB communication events (1,062 observations, roughly equally distributed across the various years); speeches by the ECB president (131 observations); speeches by all other Executive Board members (519 observations); and former ECB president Draghi’s “Whatever it takes” statement on 26 July 2012.

⁹The statistics for the German sample show a surprisingly large number of tweets where subjectivity is measured as zero. This is due to two reasons. First, the German dictionary has fewer relevant entries along the subjectivity dimension. Second, tweets in German often use English words, which are not recognised by the German dictionary. However, measuring subjectivity with the English dictionary yields qualitatively similar results.

¹⁰Note that users who tweet in German and English would be counted in both groups.

2.4 Differentiating Experts from Non-Experts

This section describes how experts are distinguished from non-experts, and how the two groups differ in their Twitter activity and their views about the ECB. In doing so, it is important to be aware that Twitter users are not representative of the entire population.¹¹ The collection of tweets about the ECB is even less likely to be representative of the entire population—the sample only contains users who publicly tweet about the ECB.

Differentiating experts from non-experts is not straightforward. Institutional Twitter accounts could be one option to identify experts, as many of these are run by economic or financial professionals, or by media. However, this identification might be too noisy—on the one hand, not all experts have institutional accounts; on the other hand, not all institutional accounts are experts in central banking or monetary policy matters. This means that experts and non-experts need to be differentiated based on their behaviour. This is done based on two main criteria.

First, it is assumed that experts are “regulars”, meaning that they often comment on ECB policies. The obvious point in time when one should expect experts to voice their opinion is on days when the monetary policy decisions are announced and commented upon in a press conference. The benchmark definition of experts therefore assumes that they comment on ECB decisions at least every second press conference. It is not required that they issue a tweet for every single press conference, in order to allow for the possibility that not every press conference is equally newsworthy, or that the experts are taking time off.

A second criterion used in the identification is the ECB centrality of the various accounts. In particular, it is assumed that non-experts write tweets about a variety of issues, and only occasionally tweet about the ECB or its policies.¹² While ECB centrality might be a good criterion to identify non-experts, it is not included in the benchmark definition of experts, for the following reason: Twitter accounts from media outlets tend to release tweets about a large range of issues, and therefore have a low level of ECB centrality. Still, tweets about the ECB issued from these accounts are most likely written by experts.

Based on these considerations, the benchmark (bm) definitions for experts and non-experts are:

$$E_i^{\text{bm}} = \begin{cases} 1 & \text{if } A_{PC,i} \geq 0.5 \\ 0 & \text{else} \end{cases} \quad (2.1)$$

$$N_i^{\text{bm}} = \begin{cases} 1 & \text{if } A_{PC,i} < 0.5 \ \& \ S_{ECB,i} < P_{25}(S_{ECB}) \\ 0 & \text{else} \end{cases} \quad (2.2)$$

where i denotes the account and $A_{PC,i}$ is the share of press conferences for which an ECB-related tweet is observed on the same day. $S_{ECB,i}$ is the share of ECB-related tweets in

¹¹Wojcik and Hughes (2019) show that U.S. Twitter users are younger, more likely to identify as Democrats, more highly educated and have higher incomes than U.S. adults overall. At the same time, there are no particular differences with regard to gender or ethnicity.

¹²Recall that the sample only includes Twitter accounts that have issued at least 100 tweets. Otherwise, there could be some accounts with a very small number of tweets, leading to extreme values of ECB centrality.

the total number of tweets originating from the account,¹³ and $P_{25}(S_{ECB})$ denotes the 25th percentile of ECB centrality across all accounts in the sample.

It is important to note that these definitions split the sample of accounts into three parts—experts and non-experts, but also a third group in between (i.e. those who did not tweet on at least every second press conference day, while having higher ECB centrality than non-experts). Effectively, this means that a (potentially large) number of accounts is discarded to better differentiate the two groups and to provide cleaner evidence on their respective behaviour.

To test for robustness, the expert and non-expert groups are redefined in various ways: for experts, a less narrow definition characterises anyone as expert who comments on at least every third press conference ($E_i^{0.33}$), a more narrow definition requires experts to comment on at least three out of four press conferences ($E_i^{0.75}$), and another alternative defines experts according to the benchmark definition (a tweet at least every second press conference), but furthermore requires a high level of ECB centrality, by only including accounts which are at least at the 75th percentile of ECB centrality across all accounts in the sample ($E_i^{centric}$).

Robustness for non-experts is tested using two variants. A less restrictive definition removes the ECB centrality criterion, and as such only requires that an account does not follow the press conference regularly ($N_i^{non-centric}$). This definition comprises all accounts that are not classified as experts in the benchmark definition. A more restrictive definition requires—in addition to the low centrality—that non-experts have few followers, i.e. are below the 25th percentile of accounts according to the number of followers ($N_i^{few\ followers}$). The idea here is to capture non-experts from the general public, rather than, e.g., politicians or experts from other fields who have many followers and occasionally comment on the ECB.

Table 2.1 provides an overview of various characteristics of experts and non-experts, according to the benchmark definition (Appendix Table B.4 shows this for the robustness definitions). Out of 255,336 accounts, roughly 24% are classified as non-experts, and around 0.5% as experts. Hence, the classification is rather conservative: 75% of accounts are discarded in order to increase the likelihood that the accounts are appropriately classified. Ratios in the German sample are similar, with 24% of the 14,780 accounts classified as non-experts and 0.1% as experts.

Given their different activity, the two groups contribute in very different ways to the overall Twitter volume. While representing around 24% of the accounts, non-experts issued only around 4% of all ECB-related tweets, whereas the 0.5% of experts contributed nearly 24%. In the German sample, 7% (8%) of tweets were issued by non-experts (experts).

Table 2.1 shows that non-experts are considerably more likely to tweet during weekends—18% of their tweets are published on Saturdays and Sundays, compared to 7% for the experts. The pattern is very similar for the accounts in German. This is in line with the notion that experts' Twitter activity takes place during business hours, and suggests that the differentiation of experts and non-experts has worked well.

There is no difference with regard to the number of followers that experts and non-experts

¹³The total number of tweets originating from a given account is observed since the creation of the account, and the number of ECB-related tweets since 2012. For accounts created before 2012, the total number of tweets since 2012 is approximated via the average number of tweets per year.

Table 2.1. Summary Statistics For Different Account Types

	<i>English Sample</i>		<i>German Sample</i>	
	<i>Experts</i>	<i>Non-experts</i>	<i>Experts</i>	<i>Non-experts</i>
Account characteristics				
Number of accounts	1,158	61,278	18	3,548
Average weekend activity	0.0691	0.1837 ***	0.0567	0.2026 *
Average percentile followers	68	68	67	66
Average percentile ECB centricity	84	12 ***	85	12 ***
Subjectivity				
Average	0.2472	0.2750 ***	0.0603	0.0211
Average of account-specific standard deviation	0.2578	0.2163 ***	0.1635	0.0335 ***
Standard deviation of account-specific average	0.0926	0.2722 ***	0.1129	0.1250
Favourableness				
Average	0.0428	0.0515	0.0264	0.0274
Average of account-specific standard deviation	0.1709	0.1498 ***	0.4242	0.2054 ***
Standard deviation of account-specific average	0.0567	0.2161 ***	0.2099	0.3354
Absolute favourableness				
Average	0.0997	0.1350 ***	0.2612	0.1576
Average of account-specific standard deviation	0.1483	0.1285 ***	0.3635	0.1802 ***
Standard deviation of account-specific average	0.0511	0.1837 ***	0.1394	0.3062 ***

Notes: The table shows summary statistics for the different account types, defined according to the benchmark definitions. Statistics related to tweet sentiment exclude any tweets issued on press conference days. Stars correspond to the following p-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Statistical significance is based on Levene's (1960) robust test statistic for the equality of variances for the standard deviation of averages, and on mean comparison tests otherwise.

have (both in the English and the German sample). Non-experts might therefore be equally influential in shaping the public discourse about the ECB as experts; understanding their behaviour is therefore of interest to the central bank.

To compare the sentiment of tweets across the two groups, tweets issued on press conference days are excluded, given that an increased frequency of tweets on those days is used to classify experts. Table 2.1 shows that mean subjectivity is significantly higher for non-experts, in line with the idea that experts provide, on average, more factual information. At the same time, looking at the within-account standard deviation of subjectivity, subjectivity is significantly more dispersed for the experts than for the non-experts. This pattern is evident for the English and the German tweets, even though statistical significance is (expectedly) less pronounced in the smaller German sample. This implies that experts issue a mixture of more factual and more subjective tweets, whereas there is less such variation in the Twitter behaviour of non-experts. Another interesting feature is that the distribution of subjectivity across accounts has a much higher standard deviation for non-experts than for experts, suggesting that the range of views expressed by non-experts is much larger.

Turning to favourableness, very similar results are obtained. Non-experts show less variation over time for a given account than experts, but there is much more variation across accounts than for experts. In addition, the strength of views that get expressed (measured via absolute favourableness i.e. opinionatedness) is higher for non-experts.

The picture that emerges therefore is that non-experts express stronger opinions, are more subjective in their views, and represent a much larger variety of views than the experts in the sample. These findings are intuitive and provide more evidence that the differentiation of accounts has indeed succeeded in singling out experts and non-experts.¹⁴

2.5 Determinants of Retweets, Likes and Replies

The analysis starts by investigating which tweets got liked, retweeted and replied to. Of the 3.2 million original tweets in English, less than 410,000 got retweeted, around 360,000 got liked and around 230,000 received a reply. These patterns are comparable in the German sample—of the 97,000 tweets, around 12,500 got retweeted, 11,300 were liked and 6,400 got replied to. Conditional on getting retweeted, liked or replied, this happens on average 4.1, 3.6 and 1.6 times in the English sample, and 2.6, 2.9 and 1.6 times in the German sample.

Retweets, likes and replies do not necessarily overlap much—in the English sample, around 201,000 tweets got retweeted but were not liked, around 152,000 tweets are liked but were not retweeted, and around 103,000 tweets were replied to but neither liked nor retweeted. This suggests that likes, retweets and replies are different concepts, requiring separate analysis.

The focus of the analysis will be how the semantic content of the original tweet affects the likelihood of being retweeted, liked or replied to - whether more factual or more subjective tweets experience more action, whether there is a “negativity bias”, and to what extent it matters how strong the views are that get expressed. These hypotheses go back to Mullainathan and Shleifer (2005), who show that newspapers slant stories toward extreme positions and to Naveed et al. (2011), who report that negative tweets are more likely to be retweeted.

These hypotheses are tested with two types of regressions. The first explains whether a tweet gets retweeted, liked or replied to, based on probit models:

$$R_j = \begin{cases} 1 & \text{if } Y_j^* > 0 \\ 0 & \text{otherwise} \end{cases} \quad (2.3)$$

$$Y_j^* = \mu_t + \beta_p p_i + \beta_l l_j + \beta_n D_j^- + \beta_f |f_j| + \beta_s s_j + \beta_N D_j^N + \beta_E D_j^E + \varepsilon_j, \quad (2.4)$$

$$\mu_t = \alpha_d + \alpha_m + \alpha_h + \alpha_t t + \alpha_{t^2} t^2, \quad (2.5)$$

where R_j denotes the dependent variable for tweet j . μ_t contains a set of control variables, namely α_d and α_m for day of week and month of year effects, α_h a dummy variable for holidays,¹⁵ and t and t^2 a linear and quadratic time trend, respectively. p_i is the percentile at

¹⁴The four most pronounced differences (in subjectivity, absolute favourableness, the standard deviation of average favourableness and weekend activity) are remarkably robust to changing the definitions of experts and non-experts (see Appendix Table B.4).

¹⁵These cover New Year (January 01), Easter (Good Friday to Easter Monday), Labour Day (May 01), Robert Schumann Day (May 09), Ascension Day, Whit Monday, Corpus Christi, Day of German

which the account i is located in the distribution of followers across all accounts—the more followers an account has, the more likely its tweets get read, liked and retweeted. l_j denotes the length of the tweet, measured by the number of characters (in 100s). The variables of interest are D_j^- , a dummy for tweets with negative favourability, $|f_j|$, the absolute value of the tweet's favourableness, s_j , its subjectivity, and two dummy variables D_j^N and D_j^E , for tweets by non-experts or by experts.

The second regression equation looks at how often a tweet gets retweeted, liked or replied to (n_j), conditional on being retweeted, liked or replied to at least once. For this analysis, the log of the corresponding number is explained using standard ordinary least squares. The explanatory variables are identical to those in the probit regression, therefore the specification can be represented by Equation (2.4), with $\ln(n_j)$ as the dependent variable.

For each of these regressions, robust standard errors are calculated. Table 2.2 reports the corresponding results; for the probit models, the table reports marginal effects.

Table 2.2. Determinants of Retweets, Likes and Replies

	English						German					
	Retweet		Like		Reply		Retweet		Like		Reply	
	Probit	OLS	Probit	OLS	Probit	OLS	Probit	OLS	Probit	OLS	Probit	OLS
Negative sentiment	0.001** (0.000)	0.002 (0.004)	0.002*** (0.000)	-0.015*** (0.004)	0.009*** (0.000)	0.004 (0.003)	0.008** (0.004)	0.061*** (0.021)	0.002 (0.003)	0.010 (0.023)	0.005* (0.003)	0.017 (0.019)
Abs(favourableness)	0.027*** (0.001)	0.053*** (0.010)	0.045*** (0.001)	0.114*** (0.009)	0.027*** (0.001)	0.015** (0.006)	0.004 (0.003)	-0.031 (0.019)	0.014*** (0.003)	0.026 (0.021)	0.010*** (0.002)	0.016 (0.018)
Subjectivity	0.015*** (0.001)	0.006 (0.006)	0.023*** (0.001)	0.014** (0.006)	0.037*** (0.001)	0.006 (0.004)	0.011* (0.006)	-0.034 (0.034)	0.025*** (0.005)	-0.009 (0.037)	0.020*** (0.004)	0.045 (0.036)
Percentile Followers	0.005*** (0.000)	0.013*** (0.000)	0.003*** (0.000)	0.010*** (0.000)	0.002*** (0.000)	0.004*** (0.000)	0.004*** (0.000)	0.010*** (0.000)	0.003*** (0.000)	0.009*** (0.000)	0.002*** (0.000)	0.006*** (0.000)
Non-expert	-0.049*** (0.001)	-0.109*** (0.008)	-0.021*** (0.001)	-0.072*** (0.007)	0.006*** (0.001)	-0.052*** (0.004)	-0.055*** (0.004)	0.099*** (0.034)	-0.010** (0.004)	0.111*** (0.032)	-0.001 (0.003)	-0.016 (0.022)
Expert	0.033*** (0.000)	0.174*** (0.003)	0.006*** (0.000)	0.085*** (0.003)	-0.010*** (0.000)	0.039*** (0.003)	-0.129*** (0.004)	-0.407*** (0.022)	-0.128*** (0.004)	-0.437*** (0.022)	-0.126*** (0.006)	-0.192*** (0.038)
No. of Characters	0.032*** (0.000)	0.159*** (0.004)	0.032*** (0.000)	0.128*** (0.003)	0.034*** (0.000)	0.034*** (0.002)	0.039*** (0.003)	0.203*** (0.018)	0.044*** (0.003)	0.189*** (0.018)	0.027*** (0.002)	0.034** (0.015)
Observations	3,219,093	409,159	3,219,093	359,706	3,219,093	231,338	96,947	12,514	96,947	11,331	96,947	6,430
R-squared		0.099		0.114		0.051		0.101		0.125		0.091

Notes: The table shows coefficient estimates for the determinants of retweets, likes and replies. Left panel: English sample; right panel: German sample. Results for probit models are marginal effects. The OLS models explain the log of the number of retweets, likes and replies, and are thus conditional on a tweet being retweeted, liked or replied to, respectively. The models control for day of week, month of year and holiday effects and allow for a linear and quadratic time trend. Numbers in brackets are standard errors. Stars correspond to the following p-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

For both languages, it is consistently found that tweets from accounts with more followers as well as longer tweets have a considerably higher likelihood of getting retweeted, liked or replied to, and conditional on being acted upon, much more often so. In addition, there is considerable seasonality (not shown in the table for brevity), both over the year and over the weekdays, as well as evidence for holiday effects and time trends.

The origin of a tweet also matters. Tweets from experts are more, tweets from non-experts less likely to be retweeted and liked than those from the bulk of the accounts. For replies, the picture is somewhat mixed, with experts' tweets being less likely to receive a reply,

Unity (October 03), All Saints' Day (November 01), Christmas (December 24, 25 and 26) and New Year's Eve (December 31).

but conditional on being replied, receiving more of them, whereas non-experts' tweets conditionally and unconditionally receive fewer replies.

Regarding the semantic content, all three sentiment measures seem to matter. First, there is evidence of a negativity bias. Tweets with a negative sentiment are more likely to be liked, retweeted and replied to (even though the conditional effect is ambiguous).¹⁶

Second, strong views are much more likely to generate retweets and likes, both unconditionally and conditionally. These effects are not only statistically significant, they are also economically large. If absolute favourableness increases from 0 to 1 (i.e. from the lowest possible to the largest possible value), the likelihood that a tweet gets retweeted (liked/replied to) increases by 2.7 (4.5/2.7) percentage points in the English sample. Such increases are sizable, given that the unconditional probability of being retweeted or liked is around 11-12% for likes and retweets, and 7% for replies.

Third, the likelihood of being retweeted or liked is also increasing in the subjectivity of the tweet, in non-negligible magnitudes. These results suggest that the discussion about the ECB on Twitter is disproportionately influenced by negative views, views that are expressed in strong language and by relatively subjective tweets—patterns that the ECB should be aware of, as such views are likely to shape the tone of the public discourse.

2.6 Responsiveness of Tweets to ECB Communication

The responsiveness of tweets to ECB communication is studied using daily data (including Saturdays and Sundays), yielding 2,532 observations.¹⁷ The reason for using aggregated data is that this allows studying not only the content of tweets, but also the amount of Twitter traffic, the number of accounts that participate in the discussion and the heterogeneity of the views that get expressed in response to the ECB's communication, i.e. variables that are most conveniently analysed at an aggregated level.

2.6.1 Hypotheses and Specification of the Econometric Model

It is first analysed to what extent Twitter traffic volume responds to ECB communication, by studying the log number of tweets posted each day. If the ECB's communication manages to reach out to non-experts, one should expect an increase in traffic not only for experts, but also for non-experts. Second, it is studied to what extent participation increases—higher volumes could result from the same users tweeting more, or from users who normally participate little or not at all joining the discussion more actively, or both. This analysis of participation is conducted using the Herfindahl-Hirschman indicator.¹⁸ Presumably, if the

¹⁶Including account fixed effects, the results become generally smaller and less significant, suggesting that the likelihood of a retweet or like does not increase if a tweet is relatively more negative, subjective or opinionated than the typical tweet from the account; rather, it is tweets from accounts that tend to write negative, opinionated or subjective tweets that get retweeted or liked more. Effects for the likelihood of getting replies are mostly unaffected by the inclusion of account fixed effects.

¹⁷All time-relevant calculations are based on Central European (Summer) Time (CET or CEST), the time in Frankfurt, Germany, the location of the ECB's headquarters.

¹⁸The indicator provides a measure of concentration—the larger the indicator, the larger the “market share” of the participating accounts on a given day. It is measured as $HHI_t = \sum_{i=1}^{U_t} s_{i,t}^2$,

ECB's communication manages to reach out to non-experts, participation should increase.

Subsequently, it is analysed whether the content of tweets changes. Their subjectivity, favourableness and absolute favourableness is studied, each time looking at the average for a given day and the daily standard deviation across tweets. The daily averages allow assessing whether the discussion on Twitter becomes more or less factual, more or less favourable and is expressed in more or less strong language. Studying the standard deviation of these variables across tweets indicates whether the views become more similar or more divergent.

These tests will allow differentiating whether Twitter is mainly a channel for information transmission or whether it hosts controversial discussions. If tweets are predominantly about relaying information, they should be rather factual (i.e. subjectivity should decline), written in relatively neutral language (i.e. absolute favourableness should fall), and be so more homogeneously across tweets (reducing the standard deviation of sentiment). Also, once the information is relayed, Twitter traffic is expected to return back to "normal" rather quickly. In contrast, a controversial discussion implies heterogeneous views across tweets, possibly for a longer time period, and is likely characterised by less factual tweets using stronger language.

For all dependent variables, identical regression equations are employed, of the form

$$x_t = \alpha_d + \alpha_m + \alpha_h + \alpha_t t + \alpha_{t^2} t^2 + \beta_{c,l}^e C_{t,l}^e + \varepsilon_t, \quad (2.6)$$

where the variables of interest are $C_{t,l}^e$, which cover the various ECB communication events e , which are allowed different lags l .¹⁹ In addition, leads are allowed for the ECB press conference, as Twitter activity intensifies already several days ahead of the press conferences.²⁰

The equations are estimated using ordinary least squares, with robust standard errors. To get at a parsimonious model specification, an increasing number of lags (and leads for the press conference) is allowed, as long these are statistically significant. This process is followed for the first specification that explains the daily number of tweets originating from all accounts. For comparability, the same lead and lag structure is then applied across all other specifications.

The first result to note, therefore, relates to the number of leads and lags that are required. With two exceptions, the various ECB communication events affect Twitter volume only on the same day. The exceptions are the ECB press conference and the "Whatever it takes" statement. For the press conference, 5 leads and 3 lags are included.²¹ For the German

where $s_{i,t}$ is the "market share" of a tweeting user i in the "tweet market" on day t ($s_{i,t} = \frac{\sum_{\text{tweets } i,t}}{\sum_{\text{tweets } t}}$), and U_t is the number of users on day t .

¹⁹We interpret the regression coefficients as causal. In line with the event study literature, it is assumed that on days of ECB communications, these are the dominant drivers of ECB-related Twitter traffic. It is unlikely that there is systematic arrival of other relevant pieces of news on these days. Reverse causality is also unlikely because the ECB communication events are largely predetermined.

²⁰This is in line with a substantial literature on financial market effects prior to the announcement of monetary policy decisions (see, e.g., Lucca and Moench (2015)).

²¹The estimated coefficients are broadly increasing over time in the up-run to the press conference, and then decline subsequently, suggesting that the effect is not triggered by other communications

sample, the time span is considerably shorter—one lead and two lags are sufficient to capture the dynamics around press conferences. The “Whatever it takes” statement requires 20 lags in English. While only 15 lags are included in the specification for the German tweets, it is apparent that this statement has led to an extraordinarily persistent discussion on the social media.

The discussion of the empirical results will first focus on the English sample, and then provide a comparison with the findings in the German sample.²²

2.6.2 Twitter Traffic and User Concentration

Table 2.3 reports the coefficient estimates for the log number of tweets and the user concentration measure. For brevity, estimates for leads and lags of the press conference and “Whatever it takes” are omitted, and the overall sum of coefficients across all lags and leads is reported in panel B. The omitted coefficients are provided in Appendix Table B.5.

Starting with the results for Twitter volume from all accounts, a reaction to all events on the same day is apparent. The press conference and “Whatever it takes” stand out in terms of magnitude, not only because they affect Twitter volume over several days, but also because of the strength of the effect on the same day: Twitter volume rises by a factor of 2 to 2.5, whereas volume for all other events increases by a factor of 0.2 to 0.5. The results for the press conference are in line with it constituting the main communication vehicle, which is also reflected in substantial media coverage, both for the ECB and the Federal Reserve (Berger et al., 2011; Binder, 2017). The responsiveness to speeches by the ECB president is around 75% higher than to speeches by the other Executive Board members, echoing earlier findings that these are more key for gauging the future path of monetary policy (e.g., Bennani et al. (2020)).

The overall response to the press conference and “Whatever it takes” (reported in panel B) demonstrates how powerful these communication events are. For the ECB press conference, Twitter volume is on average 60% higher than on normal days, for 9 consecutive days. “Whatever it takes” has been even more influential—on average, Twitter activity about the ECB was more than 140% higher than normal, for 16 consecutive days.

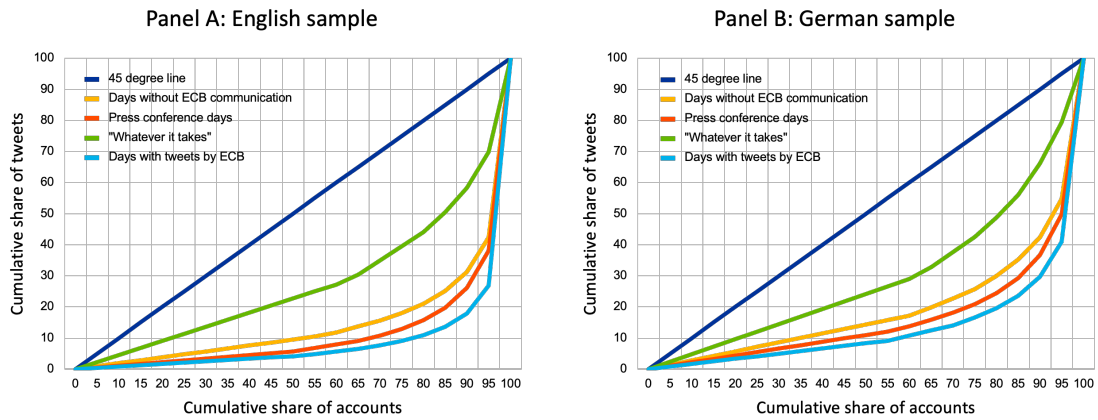
Turning to the concentration measure (reported in the right panel of Table 2.3), most events reduce concentration, i.e. the participation in the discussion is more evenly spread across accounts. The effect is sizeable. The mean concentration measure for the overall English sample is 0.006, with a standard deviation of 0.007. The drop in concentration on the ECB press conference days therefore amounts to around 60% of a standard deviation. In line with this finding, Figure 2.2 shows Lorenz curves for Twitter activity on event days and days without ECB communications, illustrating that many more people participated in the debate following “Whatever it takes” statement.

Comparing non-experts with experts yields a number of interesting insights. First, as shown in Appendix Table B.5, non-experts are not talking about upcoming press conferences more than one day ahead. Second, non-experts are not responsive to most of the more specialised communications, such as the Economic Bulletin or speeches by other Executive Board

such as speeches which breach the ECB’s quiet period (Gnan and Rieder, 2023).

²²The results are generally robust to using the different ways of classifying non-experts and experts. Details can be found in appendix Tables B.6-B.13).

Figure 2.2. Lorenz Curve of Twitter Activity by Events



Notes: The figure shows in the left (right) panel the Lorenz curve of ECB-related Twitter activity for the English (German) sample. The blue line represents the 45-degree line (which represents the line of equality). Yellow line: distribution of tweets on days without any ECB communication; red line: press conference days; green line: President Draghi’s “Whatever it takes” statement; light blue line: days with tweets by the official ECB account (and no other official communication).

members than the ECB president. Third, where they are responsive, the magnitude of the response is typically much smaller than for experts (typically, it is in the order of 60% to 70% of the experts’ response). The lower responsiveness of non-experts is also reflected by the R² of the regression models—around 63% of the variation is explained in the expert sample, as opposed to roughly 32% in the non-expert sample.

The striking exception to this difference in responsiveness is “Whatever it takes”—here, the response coefficients of experts and non-experts are very similar in magnitude. Also, Twitter traffic still shows a significant response after 20 days, for experts and non-experts alike. This suggests that “Whatever it takes” had a long-lasting effect on both groups. These findings imply that the ECB’s communication manages to reach out to experts and non-experts. They also provide a first indication that there might be different types of ECB communication events. On the one hand, some events lead to increased traffic only on the same day, with non-experts responding less than experts. On the other hand, some events lead to persistent Twitter traffic by experts and non-experts alike. These patterns are compatible with a more prominent information transmission role in the first case, and with Twitter being used as a platform for more controversial discussions (which are not resolved within one day and attract more participation by non-experts). To get a clearer picture of this dual role, it will be insightful to now look at the content of tweets, covering their subjectivity, their favourableness and their absolute level of favourableness.

2.6.3 The Content of Tweets

Table 2.4 contains the results for subjectivity, both for the daily average (left panel) and the daily standard deviation across tweets (right panel). Compared to the results reported in Table 2.3, the number of coefficients that are estimated to be statistically significant is much smaller, meaning that subjectivity is not nearly as responsive to ECB communication

Table 2.3. Explaining Twitter Traffic and Concentration of Users

	Log number of tweets						Concentration index					
	English sample			German sample			English sample			German sample		
	All	Non-experts	Experts	All	Non-experts	Experts	All	Non-experts	Experts	All	Non-experts	Experts
Panel A: Contemporaneous response												
Press Conference	2.467*** (0.083)	2.056*** (0.121)	2.816*** (0.086)	2.422*** (0.126)	1.200*** (0.181)	2.605*** (0.148)	-0.004*** (0.001)	-0.043*** (0.003)	-0.024*** (0.003)	-0.134*** (0.016)	-0.403*** (0.047)	-0.538*** (0.041)
Whatever it takes	1.975*** (0.080)	1.929*** (0.103)	1.831*** (0.090)	3.198*** (0.133)	1.530*** (0.180)	2.453*** (0.159)	-0.002*** (0.000)	-0.017*** (0.004)	-0.013*** (0.004)	-0.101*** (0.019)	-0.430*** (0.059)	-0.419*** (0.053)
Economic Bulletin	0.205** (0.089)	0.180 (0.109)	0.275*** (0.095)	-0.203 (0.141)	-0.264 (0.186)	-0.171 (0.162)	-0.001 (0.001)	-0.006* (0.004)	-0.008*** (0.003)	-0.000 (0.025)	0.067 (0.073)	-0.014 (0.060)
Accounts	0.542*** (0.086)	0.312*** (0.098)	0.853*** (0.095)	0.115 (0.149)	0.067 (0.214)	-0.146 (0.186)	-0.002*** (0.001)	-0.021*** (0.004)	-0.017*** (0.004)	-0.035 (0.024)	-0.048 (0.075)	0.005 (0.064)
Speeches by others	0.252*** (0.045)	0.084 (0.057)	0.394*** (0.053)	0.164** (0.076)	0.031 (0.097)	0.044 (0.093)	-0.002*** (0.001)	-0.006** (0.003)	-0.015*** (0.003)	-0.040*** (0.014)	-0.009 (0.034)	-0.031 (0.034)
Speeches by president	0.449*** (0.054)	0.360*** (0.074)	0.509*** (0.057)	0.863*** (0.095)	0.488*** (0.132)	1.119*** (0.104)	-0.001*** (0.000)	-0.013*** (0.003)	-0.001 (0.001)	-0.057*** (0.009)	-0.183*** (0.040)	-0.305*** (0.030)
Tweet	0.188*** (0.044)	0.150*** (0.054)	0.247*** (0.053)	0.108 (0.071)	0.076 (0.087)	0.075 (0.087)	-0.001** (0.001)	-0.008*** (0.003)	-0.012*** (0.003)	-0.032** (0.014)	-0.040 (0.031)	-0.051 (0.032)
Panel B: Overall response												
Press Conference	5.447	4.116	6.358	4.512	2.036	4.295	-0.023	-0.138	-0.226	-0.358	-0.776	-1.061
Std. error	0.269	0.335	0.315	0.256	0.350	0.308	0.003	0.016	0.023	0.034	0.112	0.109
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Whatever it takes	30.044	27.876	26.597	43.845	27.471	15.494	-0.095	-0.720	-0.875	-3.115	-6.962	-3.911
Std. error	1.059	1.052	1.276	1.230	1.518	1.008	0.011	0.081	0.099	0.212	0.516	0.354
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	2,532	2,532	2,532	2,523	1,461	1,177	2,532	2,532	2,532	2,523	1,461	1,177
R-squared	0.582	0.324	0.628	0.352	0.177	0.442	0.259	0.198	0.394	0.163	0.167	0.277
Mean(dependent var)	7.051	3.685	5.756	3.155	1.003	1.220	0.006	0.051	0.042	0.150	0.608	0.604
Stddev(dependent var)	0.905	0.880	1.158	1.192	1.037	1.115	0.007	0.047	0.068	0.182	0.359	0.339

Notes: The table shows coefficient estimates for the effect of ECB communication events on log number of tweets and the user concentration index, based on Equation (2.6). The models control for day of week, month of year and holiday effects and allow for a linear and quadratic time trend. They contain 5 (1) leads and 3 (2) lags for the press conference and 20 (15) lags for “Whatever it takes” in the English (German) sample. Panel B reports the accumulated response coefficients over all leads and lags, with coefficients that are statistically significant at least at the 10% level printed in bold. Numbers in brackets are standard errors. Stars correspond to the following p-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

events as Twitter volume.²³ Still, there are several events where subjectivity is affected, namely for the press conference, the Economic Bulletin, the accounts and the speeches by the president. In all cases, subjectivity declines, meaning that the tweets become more factual. This implies that ECB communication events lead to a more factual discussion about the ECB on Twitter—not only because of a compositional effect (whereby more Twitter users join the discussion who write more factual tweets) but also because the views of given Twitter users become more factual.²⁴ This is particularly the case for the non-experts whose distribution of subjectivity becomes narrower around a lower mean.

Table 2.5 reports the results for favourableness. There is little evidence that the ECB communication events affect mean favourableness. This was expected, as it is unlikely that all events (e.g., all speeches by the president, or all press conferences) affect public opinion

²³This can probably be explained by the fact that there is a tendency toward zero for most sentiment measures (also induced by short text), suggesting that detecting a response in sentiment is inherently difficult.

²⁴This is tested using the underlying microdata, by studying the effect of communication events in a regression with and without account fixed effects (results not reported for brevity). The estimated coefficients tend to have the same sign and remain statistically significant, which implies that the results cannot be explained by a compositional effect alone.

Table 2.4. Explaining Subjectivity

	Average subjectivity						Standard deviation of subjectivity					
	English sample			German sample			English sample			German sample		
	All	Non-experts	Experts	All	Non-experts	Experts	All	Non-experts	Experts	All	Non-experts	Experts
Panel A: Contemporaneous response												
Press Conference	-0.010*	-0.031***	0.012*	-0.012*	-0.031*	-0.023	-0.011***	-0.022***	0.002	0.028**	0.008	0.069***
	(0.005)	(0.007)	(0.006)	(0.006)	(0.017)	(0.022)	(0.003)	(0.004)	(0.004)	(0.013)	(0.017)	(0.019)
Whatever it takes	-0.003	-0.021**	0.016**	-0.007	0.040*	-0.001	-0.003	-0.013***	0.006	0.011	0.204***	0.157***
	(0.006)	(0.008)	(0.008)	(0.007)	(0.023)	(0.029)	(0.003)	(0.004)	(0.005)	(0.015)	(0.018)	(0.016)
Economic Bulletin	-0.010*	-0.008	-0.002	-0.006	-0.020	-0.001	-0.010***	-0.010**	-0.006	-0.007	-0.036***	0.014
	(0.006)	(0.009)	(0.007)	(0.008)	(0.033)	(0.028)	(0.003)	(0.004)	(0.004)	(0.017)	(0.012)	(0.025)
Accounts	-0.026***	-0.025**	-0.015*	-0.005	-0.054***	-0.055**	-0.012***	-0.011	-0.006	0.007	-0.036***	-0.027**
	(0.006)	(0.011)	(0.007)	(0.011)	(0.017)	(0.022)	(0.004)	(0.007)	(0.006)	(0.025)	(0.012)	(0.012)
Speeches by others	0.001	-0.002	0.008	-0.005	-0.004	-0.010	-0.001	-0.002	0.004	-0.005	-0.011	-0.007
	(0.004)	(0.005)	(0.005)	(0.004)	(0.014)	(0.015)	(0.002)	(0.003)	(0.003)	(0.009)	(0.008)	(0.012)
Speeches by president	-0.007*	-0.018***	-0.005	-0.001	-0.002	-0.011	-0.005**	-0.007**	-0.004	0.026**	0.025*	0.017
	(0.004)	(0.006)	(0.005)	(0.005)	(0.015)	(0.011)	(0.002)	(0.003)	(0.003)	(0.010)	(0.013)	(0.011)
Tweet	-0.005	-0.006	-0.002	-0.004	-0.010	0.011	-0.003	-0.004	-0.001	-0.007	-0.006	0.002
	(0.003)	(0.004)	(0.005)	(0.004)	(0.013)	(0.015)	(0.002)	(0.002)	(0.003)	(0.008)	(0.008)	(0.012)
Panel B: Overall response												
Press Conference	-0.068	-0.115	0.005	-0.024	-0.008	0.017	-0.016	-0.046	0.063	0.073	0.057	0.169
Std. error	0.022	0.028	0.032	0.014	0.041	0.053	0.012	0.015	0.020	0.030	0.033	0.041
p-value	0.002	0.000	0.871	0.088	0.845	0.743	0.198	0.002	0.002	0.016	0.081	0.000
Whatever it takes	0.213	0.096	0.544	0.046	0.202	-0.146	-0.072	0.018	0.076	0.845	1.188	0.570
Std. error	0.074	0.092	0.104	0.064	0.154	0.169	0.040	0.049	0.070	0.123	0.090	0.108
p-value	0.004	0.297	0.000	0.472	0.190	0.388	0.072	0.715	0.276	0.000	0.000	0.000
Observations	2,532	2,532	2,532	2,523	1,461	1,177	2,532	2,532	2,532	2,523	1,461	1,177
R-squared	0.174	0.145	0.110	0.038	0.031	0.048	0.066	0.067	0.102	0.083	0.062	0.094
Mean(dependent var)	0.256	0.269	0.229	0.035	0.029	0.036	0.280	0.284	0.263	0.098	0.022	0.033
Stdev(dependent var)	0.051	0.067	0.072	0.068	0.127	0.142	0.027	0.034	0.045	0.121	0.088	0.100

Notes: The table shows coefficient estimates for the effect of ECB communication events on the average and standard deviation of subjectivity, based on Equation (2.6). The models control for day of week, month of year and holiday effects and allow for a linear and quadratic time trend. They contain 5 (1) leads and 3 (2) lags for the press conference and 20 (15) lags for “Whatever it takes” in the English (German) sample. Panel B reports the accumulated response coefficients over all leads and lags, with coefficients that are statistically significant at least at the 10% level printed in bold. Numbers in brackets are standard errors. Stars correspond to the following p-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

in the same direction. The results for the standard deviation of favourableness are more telling. The right panel of Table 2.5 shows that the spectrum of opinions expressed narrows considerably, for most of the event types, and for experts and non-experts alike.

The last set of results, reported in Table 2.6, studies absolute favourableness, i.e. the strength of opinions. Starting from all accounts, it is apparent that most ECB communication events lead to a moderation of views, as both the average absolute favourableness and its standard deviation get reduced significantly. This is particularly true for non-experts, where absolute favourableness drops by around one a standard deviation on the occasion of the press conference, and by around 1/2 following “Whatever it takes”. ECB communication might thus be helpful in containing the strength of views expressed by non-experts.

2.6.4 Any Difference for Tweets in German?

So far, the discussion of results has focused on English tweets. A more controversial and possibly a more negative and more subjective discussion on the ECB could be expected in the German-speaking Twitter community, given the heated public debate in Germany about the ECB’s monetary policy in the aftermath of the global financial crisis (Schnabel, 2020).

Table 2.5. Explaining Favourableness

	Average favourableness						Standard deviation of favourableness					
	English sample			German sample			English sample			German sample		
	All	Non-experts	Experts	All	Non-experts	Experts	All	Non-experts	Experts	All	Non-experts	Experts
Panel A: Contemporaneous response												
Press Conference	-0.001 (0.004)	-0.007 (0.006)	0.004 (0.004)	-0.026 (0.019)	-0.037 (0.048)	0.025 (0.049)	-0.029*** (0.004)	-0.045*** (0.005)	-0.010** (0.004)	0.077*** (0.017)	0.201*** (0.037)	0.272*** (0.030)
Whatever it takes	0.017*** (0.005)	0.008 (0.007)	0.017*** (0.005)	0.049** (0.024)	0.054 (0.051)	0.096 (0.059)	-0.008* (0.004)	-0.030*** (0.006)	0.016*** (0.005)	0.160*** (0.020)	0.309*** (0.033)	0.297*** (0.039)
Economic Bulletin	0.005 (0.004)	0.008 (0.006)	0.005 (0.004)	-0.024 (0.030)	0.001 (0.050)	0.043 (0.058)	-0.009** (0.004)	-0.003 (0.006)	-0.005 (0.005)	0.007 (0.023)	-0.022 (0.036)	0.013 (0.044)
Accounts	0.005 (0.005)	0.010 (0.007)	0.005 (0.005)	-0.026 (0.034)	0.077 (0.067)	-0.032 (0.064)	-0.019*** (0.005)	-0.017** (0.008)	-0.015*** (0.005)	-0.009 (0.027)	0.043 (0.059)	-0.044 (0.035)
Speeches by others	0.001 (0.003)	-0.001 (0.004)	0.002 (0.003)	-0.016 (0.014)	0.043 (0.032)	-0.040 (0.040)	-0.003 (0.003)	-0.005 (0.003)	0.003 (0.003)	0.015 (0.013)	0.006 (0.019)	0.024 (0.024)
Speeches by president	0.004 (0.003)	0.004 (0.004)	0.007** (0.003)	-0.001 (0.015)	-0.090*** (0.030)	0.036 (0.031)	-0.010*** (0.003)	-0.014*** (0.004)	-0.010*** (0.003)	0.038*** (0.014)	0.055** (0.025)	0.157*** (0.025)
Tweet	-0.000 (0.002)	-0.002 (0.003)	0.000 (0.003)	-0.006 (0.012)	0.046* (0.027)	-0.009 (0.039)	-0.003 (0.002)	-0.004 (0.003)	-0.003 (0.003)	0.024** (0.011)	0.013 (0.016)	0.009 (0.022)
Panel B: Overall response												
Press Conference	-0.010	0.003	0.025	-0.047	-0.171	-0.052	-0.100	-0.089	-0.005	0.169	0.304	0.575
Std. error	0.017	0.021	0.020	0.049	0.102	0.134	0.016	0.020	0.022	0.036	0.066	0.081
p-value	0.563	0.888	0.219	0.337	0.094	0.701	0.000	0.000	0.811	0.000	0.000	0.000
Whatever it takes	-0.084	-0.197	0.021	-0.952	-0.512	-1.905	0.179	0.005	0.427	1.700	2.581	1.676
Std. error	0.050	0.066	0.063	0.225	0.399	0.466	0.050	0.062	0.067	0.190	0.259	0.257
p-value	0.095	0.003	0.736	0.000	0.199	0.000	0.000	0.932	0.000	0.000	0.000	0.000
Observations	2,532	2,532	2,532	2,523	1,461	1,177	2,532	2,532	2,532	2,523	1,461	1,177
R-squared	0.069	0.053	0.043	0.034	0.047	0.050	0.132	0.096	0.105	0.088	0.097	0.181
Mean(dependent var)	0.047	0.051	0.038	0.042	0.032	0.029	0.205	0.218	0.171	0.377	0.114	0.172
Stdev(dependent var)	0.035	0.047	0.044	0.183	0.307	0.338	0.036	0.045	0.048	0.166	0.196	0.218

Notes: The table shows coefficient estimates for the effect of ECB communication events on the average and standard deviation of favourableness, based on Equation (2.6). The models control for day of week, month of year and holiday effects and allow for a linear and quadratic time trend. They contain 5 (1) leads and 3 (2) lags for the press conference and 20 (15) lags for “Whatever it takes” in the English (German) sample. Panel B reports the accumulated response coefficients over all leads and lags, with coefficients that are statistically significant at least at the 10% level printed in bold. Numbers in brackets are standard errors. Stars correspond to the following p-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Overall, the patterns for Twitter traffic are similar in the German sample: i) there are same-day reactions to many communication events, for experts and for non-experts; ii) reactions following the press conference and “Whatever it takes” are particularly strong and more persistent; iii) non-experts tend to be generally less responsive than experts.

At the same time, there are substantial differences with regard to the sentiment of the tweets, most pronounced for “Whatever it takes”, which has been even more influential in the German-speaking Twitter community than among the English. On the day of the announcement, the number of tweets increases more than 1.5 times compared to the English sample, while the concentration measure falls much more (both in absolute and in relative terms). In addition, the overall responsiveness of the non-experts was considerably larger than the one of the experts.

“Whatever it takes” consistently raised the standard deviation of subjectivity, for experts and non-experts, both instantaneously and over the entire discussion, and by large amounts. Also the tone of the tweets is affected; German tweets, in particular those by experts, have become more negative. And the spectrum of views widened; both the standard deviation of favourableness and the standard deviation of absolute favourableness increased substantially.

Table 2.6. Explaining Absolute Favourableness

	Average absolute favourableness						Standard deviation of absolute favourableness					
	English sample			German sample			English sample			German sample		
	All	Non-experts	Experts	All	Non-experts	Experts	All	Non-experts	Experts	All	Non-experts	Experts
Panel A: Contemporaneous response												
Press Conference	-0.021*** (0.004)	-0.039*** (0.005)	-0.008** (0.003)	0.033** (0.016)	0.101** (0.043)	0.013 (0.041)	-0.023*** (0.003)	-0.035*** (0.004)	-0.009*** (0.004)	0.054*** (0.012)	0.155*** (0.031)	0.230*** (0.027)
Whatever it takes	-0.004 (0.004)	-0.021*** (0.006)	0.011*** (0.004)	0.095*** (0.021)	0.094** (0.046)	0.154*** (0.053)	-0.005 (0.003)	-0.022*** (0.005)	0.014*** (0.004)	0.112*** (0.015)	0.276*** (0.029)	0.229*** (0.032)
Economic Bulletin	-0.005 (0.004)	-0.001 (0.006)	-0.002 (0.004)	0.023 (0.026)	-0.036 (0.050)	-0.017 (0.054)	-0.008** (0.004)	-0.005 (0.005)	-0.006 (0.004)	-0.003 (0.017)	-0.019 (0.031)	-0.006 (0.036)
Accounts	-0.016*** (0.004)	-0.016** (0.007)	-0.012*** (0.004)	0.001 (0.028)	0.009 (0.065)	-0.099* (0.053)	-0.012*** (0.004)	-0.010* (0.006)	-0.011** (0.004)	-0.011 (0.019)	0.023 (0.049)	-0.035 (0.034)
Speeches by others	-0.004 (0.002)	-0.006** (0.003)	0.000 (0.003)	0.008 (0.013)	0.012 (0.029)	0.004 (0.035)	-0.003 (0.002)	-0.003 (0.003)	0.002 (0.003)	0.010 (0.010)	-0.004 (0.016)	0.019 (0.020)
Speeches by president	-0.005** (0.002)	-0.010*** (0.004)	-0.004* (0.002)	0.014 (0.014)	-0.016 (0.028)	0.016 (0.029)	-0.007*** (0.002)	-0.008** (0.003)	-0.008*** (0.003)	0.026*** (0.009)	0.052** (0.022)	0.133*** (0.020)
Tweet	-0.003 (0.002)	-0.005* (0.003)	-0.003 (0.003)	0.017 (0.011)	0.026 (0.025)	0.009 (0.033)	-0.002 (0.002)	-0.004 (0.002)	-0.002 (0.003)	0.018** (0.009)	0.006 (0.014)	0.007 (0.019)
Panel B: Overall response												
Press Conference	-0.078	-0.089	-0.030	0.061	-0.051	0.069	-0.077	-0.060	-0.002	0.146	0.269	0.494
Std. error	0.014	0.018	0.018	0.038	0.091	0.112	0.013	0.015	0.018	0.028	0.058	0.069
p-value	0.000	0.000	0.089	0.103	0.576	0.539	0.000	0.000	0.898	0.000	0.000	0.000
Whatever it takes	0.169	-0.043	0.360	-0.065	-0.059	0.153	0.048	-0.058	0.247	1.307	2.403	1.527
Std. error	0.045	0.057	0.055	0.190	0.371	0.408	0.042	0.050	0.057	0.153	0.229	0.213
p-value	0.000	0.450	0.000	0.731	0.873	0.707	0.253	0.241	0.000	0.000	0.000	0.000
Observations	2,532	2,532	2,532	2,523	1,461	1,177	2,532	2,532	2,532	2,523	1,461	1,177
R-squared	0.133	0.102	0.085	0.044	0.055	0.046	0.122	0.079	0.090	0.087	0.097	0.184
Mean(dependent var)	0.116	0.129	0.095	0.241	0.175	0.227	0.177	0.186	0.151	0.324	0.100	0.150
Stdev(dependent var)	0.031	0.042	0.039	0.159	0.280	0.290	0.029	0.035	0.040	0.126	0.167	0.185

Notes: The table shows coefficient estimates for the effect of ECB communication events on the average and standard deviation of absolute favourableness, based on Equation (2.6). The models control for day of week, month of year and holiday effects and allow for a linear and quadratic time trend. They contain 5 (1) leads and 3 (2) lags for the press conference and 20 (15) lags for “Whatever it takes” in the English (German) sample. Panel B reports the accumulated response coefficients over all leads and lags, with coefficients that are statistically significant at least at the 10% level printed in bold. Numbers in brackets are standard errors. Stars correspond to the following p-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

These findings suggest that “Whatever it takes” has not only led to a long-lasting and intense discussion by German-speaking non-experts, it has also been highly controversial and was overall met with a certain degree of negativity—a prime example of how Twitter constitutes a platform for controversial discussions about the ECB’s policies.

2.7 Conclusion

Following the global financial crisis, the subsequent use of unconventional monetary policy tools and the broadening of central bank mandates, many central banks have put more emphasis on communication with non-expert audiences. This endeavour raises several new challenges, since compared to the traditional audience, non-experts are less knowledgeable about central banking matters and might not even be reached by central bank communication. Accordingly, some commentators predicted that central banks’ attempts to communicate with the general public are bound to fail (Blinder, 2018).

Against this background, this paper uses ECB-related Twitter traffic to shed light on the question whether central banks can reach out to non-experts. It differentiates expert and non-expert Twitter accounts, which allows testing to what extent non-experts are responsive

to central bank communication, and how their views evolve around such communications.

Twitter traffic by experts and non-experts is responsive to the ECB's communication, in two ways. First, there are communication events (typically the relatively more technical ones) where Twitter mainly serves as an information transmission channel. For these events, Twitter traffic returns to normal within one day, non-experts are less involved than experts, and the views expressed tend to converge. Second, on other occasions Twitter serves as a platform for controversial discussions, which last several days, draw in many non-experts, and show a divergence of views. This has particularly been the case for former President Draghi's "Whatever it takes" statement and the ensuing discussion among German-speaking Twitter users.

A lot of the ECB-related Twitter traffic stems from retweets of tweets, implying that some opinions get shared widely, and are therefore more influential in shaping non-experts' views about the ECB. This paper shows that tweets are more likely to get retweeted, liked or replied to if they express negative views, are written in a strong language and if they are less factual.

These findings have important implications for central banks. First, they suggest that central bank communication manages to reach out to non-experts. Second, the retweet, like and reply analysis suggests that negative opinions, strong views and more subjective contributions are likely to be shared more often. At the same time, central bank communication has the potential to make discussions in social media somewhat more factual and moderate. Communication with non-experts is therefore not a road to nowhere. However, whether it ultimately succeeds in fostering trust, making central banks accountable, or influencing agents' inflation expectations and behaviour remains an open issue.

Incentivizing Inflation Expectations*

Abstract Inflation expectations are crucial for economic modeling and policymaking. Despite the well-established role of incentives in experimental economics, all major surveys of inflation expectations pay a flat participation fee. This lack of marginal incentives extends to many information provision experiments—often designed as randomized controlled trials (RCTs). We introduce marginal incentives into a standard survey of inflation expectations. Marginal incentives significantly alter expectation distributions, reducing mean forecasts, cross-sectional disagreement, and closing the gender expectations gap. Furthermore, in an embedded RCT, marginal incentives lead to greater responsiveness to information provision, contrasting with null effects under flat fees. These findings underscore the importance of marginal incentives in surveys and survey-based experiments to enhance data validity, strengthen empirical research, and better inform policymaking.

Keywords: Inflation expectations, survey methodology

JEL: E31, C83

3.1 Introduction

Macroeconomics has recently seen two significant trends: the widespread integration of survey-based beliefs into empirical research and policymaking, and the adoption of information provision experiments embedded within economic surveys. The incorporation of survey-based belief data has markedly advanced our understanding of household expectations, provided deep insights into how households form these expectations and tested long-held fundamental assumptions—rational expectations in particular—within macroeconomic theory.¹ Additionally, central banks have increasingly relied on survey-

*This chapter is based on co-authored work with Sergii Drobot, Daniela Puzzello, and Ryan Rhoades (Drobot et al., 2024). This study received ethical approval from the University of Oxford (Reference: ECONCIA23-24-04), has been pre-registered (AEA RCT Registry ID: AEARCTR-0014328), and benefited from a British Academy/Leverhulme Small Research Grant.

¹For a review of this literature, see D’Acunto and Weber (2024) and Weber et al. (2022). For canonical examples, see Coibion and Gorodnichenko (2015); Coibion et al. (2018). For an excellent

based measures of inflation expectations to inform both conventional and unconventional policy decisions, underscoring the practical importance of accurately eliciting and interpreting these beliefs. Concurrently, adopting information provision experiments, particularly randomized controlled trials (RCTs), embedded into large-scale surveys has enabled researchers to establish causal relationships between information dissemination and expectations, economic decision-making, economic outcomes, and central bank communication and expectations. These experiments typically involve providing participants with specific information or interventions and measuring the subsequent impact on their expectations and behaviors. By leveraging RCTs, economists can isolate the effects of information on expectations, thereby enhancing the robustness of their empirical findings.

Survey-based beliefs and information provision experiments almost always employ unincentivized or flat-fee incentive structures. In contrast, experimental economics has long recognized the value of marginal incentives in ensuring that participants reveal their true preferences and beliefs. The induced value theorem (Smith, 1976) posits that marginal incentives align participants' self-interest with truthful reporting, thereby enhancing data quality.² Neglecting marginal incentives may inadvertently and unnecessarily introduce measurement errors and biases into macroeconomic belief surveys, undermining the reliability of empirical conclusions drawn from such data.

In this paper, we design an experiment to test how marginal incentives—rooted in the induced value theorem Smith (1976)—affect macroeconomic beliefs elicited via survey and learning rates within a simple information provision experiment. We replicate portions of the NY Fed's Survey of Consumer Expectations (SCE) methodology for eliciting inflation expectations and then add an information intervention following common RCT practices. This means that we elicit inflation expectations twice: before the information intervention (*the prior*) and afterwards (*the posterior*). Our focus on the SCE is due to its widespread use in both academic research and policymaking (e.g., see Armantier et al. (2024), D'Acunto and Weber (2024) and Weber et al. (2022)).

Participants in our experiment are sorted into one of four treatment conditions: *Flat*, *Prior*, *Post*, and *Both*. *Flat* serves as our baseline group, where participants receive a fixed payment for participation regardless of their responses. This exactly mimics the incentive structure active in all major surveys of expectations and the majority of information provision experiments. In the *Prior* group, participants are offered marginal incentives based on the accuracy of their point forecasts for one-year-ahead inflation before any information is provided. In the *Post* group, marginal incentives are applied after participants have received the Federal Open Market Committee's (FOMC) inflation outlook, rewarding accurate probabilistic forecasts. Treatment group *Both* combines these approaches, offering incentives for both prior and posterior forecasts. We calibrate incentives so that the time-value of expected total earnings is constant across all treatments and aligns with participation remuneration in the NY Fed's SCE. This design allows us to evaluate how different incentive structures influence elicited inflation expectations and the extent to which participants update their beliefs in response to new information.

review of information provision experiments, see Haaland et al. (2023).

²There is ample evidence demonstrating the advantage of marginal incentives. For example, Nelson and Bessler (1989) and Palfrey and Wang (2009) found incentive-compatible scoring rules outperform alternatives. Gächter and Renner (2010), Wang (2011), and Trautmann and van de Kuilen (2014) showed that incentivized elicitation methods dominate unincentivized ones.

The macroeconomic literature is divided on whether incentivized elicitation improve belief accuracy. The problem of ‘cheap talk’ for elicited inflation expectations has been touched on by a few studies, raising doubt about data accuracy or reliability because respondents often lack proper economic incentives (Pesaran and Weale, 2006; Manski, 2004). For instance, Inoue et al. (2009) question the accuracy of reported inflation expectations, as they find that implicitly measuring inflation expectations through consumption data does a better job at predicting actual inflation than the reported beliefs, especially for the lower educated. Keane and Runkle (1990) question whether reported expectations are simply cheap talk or reflect actual beliefs. They find evidence for the latter—at least for the case of professional forecasters who have strong incentives to report rational and accurate expectations for reasons concerning their professional credibility and reputation. These circumstances do not directly transfer to households. Armantier et al. (2015) find a strong correlation between non-incentivized inflation expectations and investment choices in an incentivized investment experiment, except for respondents of lower education and financial literacy, suggesting overall that marginal incentives might not always be necessary. Roth and Wohlfart (2020) report no significant effect of incentives on beliefs about the likelihood of a recession. Similarly, Andre et al. (2022) find no effects for incentives on reported unemployment expectations. Pooling unemployment and inflation expectations, the authors find no significant difference between incentivized and unincentivized beliefs overall in a joint test. However, they do find that incentivizing inflation expectations shifts these moderately closer to expert forecasts. In addition, incentives increase the time taken to respond, a measure for exerted effort. Notably, Andre et al. (2022) use a clever approach to explore whether incentives affect subjective beliefs by linking rewards to second-order beliefs—participants were incentivized to match the average expert’s forecast rather than their own subjective inflation expectations. While this method provides valuable insights into how incentives might shape beliefs about expert opinion, it differs from approaches that focus on first-order beliefs, where forecasts are benchmarked against actual future outcomes.

Our approach builds on these insights but represents a significant departure from previous work by employing an incentive structure within a context closely aligned with the SCE. This ensures that the results from our treatments can be readily interpreted against a backdrop of previous studies, thus facilitating the interpretation and integration of our findings into the existing literature. Additionally, our incentive structure is both more direct and less complex. Our experiment directly incentivizes both point and probabilistic inflation forecasts, ensuring participants are motivated to provide accurate predictions and limiting the potential for confusion driven through complex incentives. Danz et al. (2022) and Abeler et al. (2023) demonstrate that complex incentive schemes can lead to misunderstandings, potentially resulting in less truthful reporting and shrouding. Additionally, their design relied on a lottery system where only a few participants were selected for payout, while we provide marginal incentives to all participants. We also directly incentivize updating in our study. This requires participants to update their beliefs after receiving new information, a crucial component that allows us to observe how marginal incentives affect not just initial beliefs but also learning and belief adjustments over time. This design is crucial for understanding how participants process and incorporate new information, something previous studies have not fully explored.

Our findings reveal that imposing marginal incentives significantly alters the distribution of

reported inflation expectations. Specifically, respondents subjected to marginal incentives provide significantly less extreme forecasts on average (mean of point expectations falls from 6.1% without marginal incentives to 2.7% with such incentives) and exhibit significantly less cross-sectional forecast disagreement (the standard deviation of point expectations drops by a third from 23.78 to 16.98). These patterns emerge regardless of whether we consider elicited priors or posteriors. Incentives further cause elicited expectations to be more consistent with professional forecasts, and resolve gender differences in expectations. In the context of RCTs, marginal incentives significantly enhance estimated learning rates, indicating that participants adjusted their beliefs more substantially and consistently in response to provided information. Estimated learning rates are significantly stronger under marginal incentives than those estimated using a traditional flat-fee scheme, suggesting that central bank forecasts might be able to coordinate and manage inflation expectations even in high inflationary environments. Finally, we find that incentives directly raise the effort that respondents exert when answering the survey, as reflected in the time they spend completing the survey.

These results have profound implications for empirical macroeconomic research and policymaking. First, they suggest that the current reliance on unincentivized survey methods may lead to biased or inaccurate measures of economic expectations, potentially skewing research findings and policy decisions. Secondly, incorporating marginal incentives into belief elicitation processes can improve the validity of survey data, leading to more reliable insights into household expectations and behaviors. Lastly, our findings highlight the necessity for macroeconomic surveys and experiments to adopt incentive mechanisms to ensure the integrity and accuracy of the data they collect or at least to take them into account. Indeed, our approach can be used to provide estimates of the proportion of the measurement error due to the absence of marginal incentives.

By bridging the gap between experimental economics and macroeconomic survey methodologies, our study underscores the critical importance of incentive structures in the collection of survey-based macroeconomic beliefs and in information provision experiments. Adopting marginal incentives not only enhances data quality but also strengthens the empirical foundations upon which economic theories and policies are built. As survey-based beliefs continue to play a pivotal role in shaping economic models and policy frameworks, ensuring their accuracy through appropriate incentive mechanisms becomes indispensable for advancing both academic research and practical policymaking.

3.2 Experimental Design

We pursue two primary objectives that shape our experimental design. First, we investigate whether and how the implementation of marginal incentives alters survey-based belief measures. Second, we examine whether marginal incentives can influence beliefs collected through a survey-based RCT, a widely adopted methodology in experimental macroeconomics. To achieve these aims, our experiment must generate reliable survey-based beliefs free from the influence of extraneous information provision while simultaneously conducting an information provision experiment.³

³The complete survey is shown in Appendix C.3. We use oTree to code the interface (Chen et al., 2016).

To address these objectives, we design an individual-choice survey that elicits both prior and posterior one-year-ahead expectations of annual inflation from each participant. Specifically, we elicit priors as point expectations (see Figures C.5-C.8) and posteriors as probabilistic forecasts (see Figures C.14-C.16). In addition to eliciting priors, we also ask for their point beliefs about inflation over the past 12 months to control for perceived inflation (these are not incentivized in any of the treatments). Between these measures, participants receive a summary of the FOMC’s most recent inflation expectations, including median forecasts for 2024 and 2025 and corresponding range forecasts (see Figure C.11). This is the information provision intervention. Additionally, we collect participants’ expectations for food and gas prices both before and after the information provision, ensuring that questions focused on inflation are adequately separated from the information provision and from each other to minimize bias (Haaland et al., 2023; Stantcheva, 2023). Importantly, we base the wording, response options, and overall survey structure on the carefully designed New York Fed’s Survey of Consumer Expectations (Armantier et al., 2024, 2017; Bruine de Bruin et al., 2010).⁴ Worth noting is that we adopted the welcoming language of the SCE intended to activate participants’ intrinsic motivation (see Figure C.2).

We implement a between-subjects design by randomizing participants into one of the four treatments, summarized in Table 3.1. Our baseline treatment, *Flat*, provides participants with a fixed fee without any marginal incentives. To match the time-value of money earned by participants in the SCE, we scale the *Flat* payment accordingly. This payment is divided into two parts: a fixed fee of \$2 paid immediately upon survey completion and an additional \$4 paid in September 2025, aligning with the forecast period. This delayed payment controls the timing of bonus payments necessary for other treatments and avoids potential selection effects.

The three other treatments introduce marginal incentives based on the accuracy of participants’ one-year-ahead inflation forecasts. In *Prior*, participants receive a bonus payment contingent on the forecast error relative to the realized annual Personal Consumption Expenditures (PCE) inflation reported by the Bureau of Economic Analysis (BEA) in September 2025. A perfect forecast earns a bonus of \$10. Each additional percentage-point (pp) forecast error reduces the bonus by half.⁵ This scoring rule is common in learning-to-forecast experiments in experimental macroeconomics.⁶ In *Post*, we pay participants $\$10 * weight_i$ where $0 \leq weight_i \leq 1$ is the probability weight assigned by the participant to bin i that contains

⁴The SCE measures U.S. consumers’ expectations on key economic variables like inflation, aiding policymakers and researchers in understanding consumer sentiment and behavior. For example, it helps the Federal Reserve assess inflation expectations, guide interest rate decisions, and forecast spending and savings trends. Its questions are also widely used in academic research.

⁵While Armantier and Treich (2013) highlight the potential for Proper Scoring Rules (PSRs) to distort beliefs when respondents have financial stakes or hedging opportunities, our inflation forecasting experiment differs in several key ways. Unlike prediction markets or controlled probabilistic events, our respondents forecast a well-known macroeconomic variable, allowing them to anchor beliefs onto experience, news, or forecasts from credible institutions. This can minimize the distortions typically associated with PSRs in more abstract or game-theoretic settings. Further, inflation forecast is fundamentally a setting of ambiguity rather than risk, and our participants lack opportunities to hedge. Additionally, incentives in our setting appear to enhance effort (see Section 3.3.5) and align forecasts more closely with professional expectations, consistent with thoughtful engagement rather than distortion.

⁶See McMahon and Rholes (2023) and Rholes and Petersen (2021) for examples.

realized inflation. For example, if inflation turns out to be 5% and a participant assigned probability weight .2 to the bin for 4% to 8%, then the participant will earn $\$10 \cdot .2 = \$2$.⁷ For *Both*, a subject faces either the point or probabilistic marginal incentive scheme with equal likelihood.

To calibrate incentives, we analyze average forecast errors using NY Fed’s one-year-ahead forecasts and actual inflation data from FRED. The average forecast error was 1.68 pp across the entire sample and 1.16 pp in the most recent six observations. Based on an estimated annual discount rate ($\beta = 0.8$) from Warner and Pleeter (2001), we set the maximum payoff for a perfect forecast so that a participant’s expected earnings in present-value terms align with the time-value for participants in the NY Fed’s SCE. For our 5-minute survey, this results in a total payout of about \$6, with 33% (\$2) allocated as a show-up fee.

In *Prior*, we apply this marginal incentive scheme to the point forecast of inflation collected before the information provision. In *Post*, the scheme is applied to probabilistic forecasts collected after the information provision. In *Both*, we inform participants we will impose marginal incentives on either the point or probabilistic forecast with equal probability, but not both.

Table 3.1. Overview of Treatments

Treatment	Prior	Posterior
Flat	Unincentivized	Unincentivized
Prior	Incentivized	Unincentivized
Post	Unincentivized	Incentivized
Both	Incentivized	Incentivized

Notes: This table shows the four treatments that differ in incentivizing elicited prior and/or posterior inflation expectations (before and after information provision). Priors are elicited using point forecast questions, while posteriors are elicited using probabilistic bin forecast questions.

3.2.1 Hypotheses

Before moving on to results, we offer two hypotheses regarding the impact of marginal incentives in our experiment. These hypotheses are grounded in the notion that financial incentives enhance cognitive effort and reduce biases in self-reported data, leading to more reliable and valid measures of economic beliefs. This logic is the basis of the longstanding practice of employing marginal incentives to discipline choice data and reduce measurement error in experimental economics (Smith, 1976; Smith and Walker, 1993; Camerer and Hogarth, 1999). In our context, we expect incentives to reduce upward bias, forecast errors, and the occurrence of outliers.⁸ Further, incentives will increase individuals’ attention to the provided information, even at a time where attention to inflation is already high, and thus

⁷Our incentive schemes are simple, though not incentive-compatible. We opt for simplicity because previous experimental studies suggest that simpler incentives can be more effective than more complex, incentive-compatible designs (e.g., Danz et al. (2022)). In a companion paper, we explore the role of incentive-compatibility in designing incentives in the context of the SCE.

⁸In the inflation expectations literature, upward bias refers to the tendency of individuals or survey respondents to systematically overestimate future inflation compared to actual inflation outcomes. This bias has been widely documented in household surveys and was also present at the time of our survey.

learning rates from information are expected to be lower (Weber et al., 2025; Mackowiak and Wiederholt, 2024).

Hypothesis 1 (Survey-Based Beliefs): *The cross-sectional distribution of inflation expectations has lower mean and lower variance with marginal incentives. Further, with marginal incentives, forecasts are closer to those of professional forecasters.*

Hypothesis 2 (Learning Rates): *In the context of the RCT, marginal incentives increase learning rates. That is, participants adjust their beliefs more substantially and consistently in response to the information provision compared to those without incentives.*

3.2.2 Data

We collect 1,000 observations—250 per treatment—from US residents via Prolific on September 14, 2024. Prolific provides information on participants’ demographic characteristics such as age, gender, income or race (see Figure C.1 in the Appendix for a comparison of demographic characteristics across treatment groups and with the sample of the SCE).⁹ The chosen sample size is based on power calculations (see Section C.2 in the Appendix). With few exceptions, we winsorize the data at the 1% and 99% levels to mitigate the impact of extreme outliers on our main results.

3.3 Results

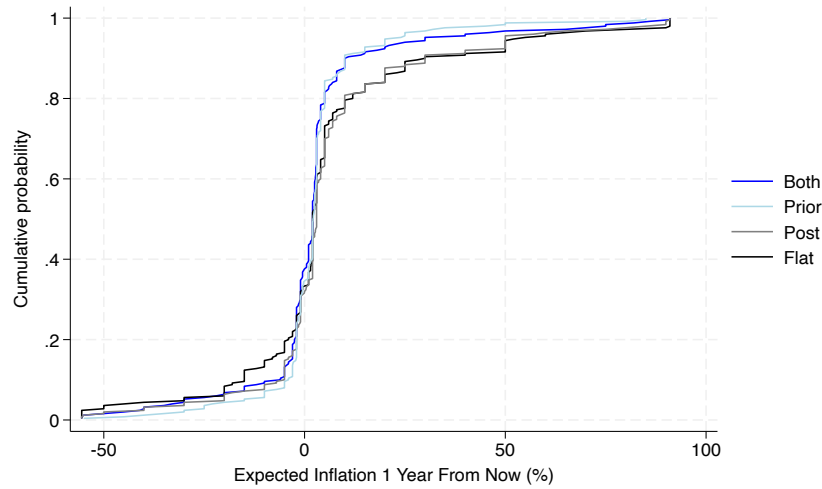
This section details the results of our survey. We first show how incentives affect elicited expectations, highlighting that incentivized expectations become more consistent with SPF forecasts, exhibit lower disagreement, raise attention to inflation, and diminish the puzzle of gendered expectations. Second, we show how incentives within an information provision experiment affect updated expectations and learning rates. Finally, we show how incentives raise participant effort.

3.3.1 The Effect of Incentives on Elicited Expectations

We first consider whether marginal incentives influence respondents’ one-year-ahead inflation expectations, which we illustrate in Figure 3.1. This figure shows the cumulative distribution functions (CDFs) of inflation expectations across the different treatment groups, expressed in percentage points. The treatments imposing marginal incentives—*Both* (blue curve) and *Prior* (yellow curve)—are contrasted with *Flat* (red curve) and *Post* (green curve), which do not include marginal incentives. The *Flat* and *Post* treatments mimic the typical approach used in all major macroeconomic surveys and thus reflect the incentive mechanism underlying the majority of belief data used in belief-based research in empirical macroeconomics.

⁹Coincidentally, there is a relatively higher proportion of females in treatments *Prior* and *Both*. Some studies have shown that females tend to have higher inflation expectations. Since we observe higher inflation expectations in the *Flat* treatment, we do not believe that this affected our results. Further, we control for gender in our regressions.

Figure 3.1. CDFs of Expected Inflation By Treatment



Notes: The Figure shows cumulative distribution functions (CDFs) of inflation expectations across the different treatment groups, expressed in percentage points. Data are winsorized at the 1% and 99% levels.

Our results show that imposing marginal incentives when eliciting inflation expectations (i.e., in *Prior* and *Both*) generate significantly different belief distributions than do flat-fee incentives (i.e., the *Flat* and *Post* treatments). More specifically, imposing marginal incentives during belief elicitation leads participants, on average, to expect significantly less extreme inflation values. These incentivized expectations appear more “reasonable” when evaluated against historical inflation data, current inflation trends, and the Fed’s contemporaneous policy stance and outlook. Put simply, the beliefs elicited under marginal incentives resemble more informed expectations.

The primary impact of these incentives manifests in the expectations of respondents who foresee inflation, rather than deflation. Under marginal incentives, respondents expecting inflation predict significantly lower price growth relative to unincentivized treatments. For those anticipating deflation, we similarly observe a muted expectation of price change under marginal incentives, suggesting that the incentives temper both inflationary and deflationary beliefs.

This distinction arises despite holding constant across treatments all other aspects of the incentives, including the timing and expected amounts of payments. We show that merely altering the structure of belief elicitation in a feasible way that imposes no additional cost relative to prevailing approaches can substantially change the nature of respondents’ reported expectations. Importantly, this change occurs without modifying participants’ perceptions of the data-generating process, introducing asymmetric information, or altering other fundamental aspects of the decision environment.

Table 3.2 summarizes the mean and standard deviation of point expectations for participants who faced marginal incentives or not. The unincentivized group has a higher mean (6.13) compared to the incentivized group (2.73), and the standard deviation is also larger in the non-incentivized group (23.78 vs. 16.98), indicating higher cross-sectional disagreement among unincentivized forecasters.

Table 3.2. Summary Statistics and Variance Comparison of Inflation Expectations

	Mean	Standard Deviation	N
Unincentivized	6.13	23.78	500
Incentivized	2.73	16.98	500
All Data	4.43	20.72	1,000

Test for Equality of Variances		
Test Type	Test Statistic	p-value
Levene's Test (Mean)	31.54	p < .001
Levene's Test (Median)	21.31	p < .001
Levene's Test (Trimmed Mean)	23.32	p < .001
F-Test (Variance Ratio)	1.9594	p < .001

Notes: This table shows mean and variances of the elicited prior belief of inflation $\mathbb{E}(\pi_{prior})$ by incentive treatments. Unincentivized is comprised of treatments *Flat* and *Posterior*, while Incentivized is comprised of *Both* and *Prior*.

We test for the equality of variance across incentive schemes using both Levene's tests and the F-test for variance ratios. All tests strongly reject the null hypothesis that the variances are equal ($p < 0.001$). Thus, imposing marginal incentives reduces the mean of point inflation expectations and leads to lower cross-sectional forecast disagreement.

To quantify the effect of incentives on extreme forecasts, we estimate the following regression:

$$|\mathbb{E}(\pi_{prior})| = \alpha + \gamma_i \text{Treatment}_i + \beta \mathbb{X} + \epsilon \quad (3.1)$$

where $i \in \text{Flat}, \text{Post}, \text{Both}, \text{Prior}$ denotes the incentive treatment groups, and $\mathbb{X}$ represents a vector of control variables. The results of this regression are displayed in Table 3.3. Column (1) provides baseline results without controls, and each subsequent column progressively introduces additional covariates.

Our hypothesis, which we detail in Section 3.2, posits that marginal incentives significantly alter the distribution of inflation expectations. The results strongly support this hypothesis. As shown in Table 3.3, the coefficients for *Both* and *Prior* indicate that marginal incentives significantly reduce extreme expectations of inflation rates, compared to the unincentivized *Flat* treatment. Specifically, respondents in the *Prior* group report significantly less extreme inflation expectations compared to those in the *Flat* treatment, with absolute forecast values approximately half as large on average ($p < 0.001$). The effect in *Both* is somewhat less pronounced but still substantial, with respondents providing significantly less extreme inflation forecasts—about a third lower on average compared to *Flat* ($p < 0.01$). These effects are robust to controlling for the expected direction of price change (column (2)), for a participant's gender (column (3)), controlling whether a subject has at least an undergraduate degree (column (4)), and controlling for a participant's one-year-ahead economic sentiment (column (5)).

The influence of incentives on expectations highlights the need for careful consideration when interpreting survey-based belief measures and the conclusions drawn from them. If belief elicitation is highly sensitive to the presence of incentives, it becomes crucial to

Table 3.3. Effects of Incentives on Inflation Expectations

	(1) $\mathbb{E}(\pi_{Prior})$	(2) $\mathbb{E}(\pi_{Prior})$	(3) $\mathbb{E}(\pi_{Prior})$	(4) $\mathbb{E}(\pi_{Prior})$	(5) $\mathbb{E}(\pi_{Prior})$
Post	-1.770 (1.916)	-1.785 (1.915)	-1.473 (1.865)	-1.314 (1.891)	-0.994 (1.832)
Both	-5.268*** (1.802)	-5.213*** (1.804)	-5.828*** (1.778)	-5.757*** (1.791)	-6.052*** (1.732)
Prior	-7.794*** (1.620)	-7.779*** (1.620)	-8.492*** (1.610)	-8.540*** (1.612)	-8.705*** (1.564)
Deflation		-1.248 (1.183)	-1.609 (1.197)	-1.609 (1.196)	-0.0966 (1.156)
Male			-7.175*** (1.147)	-7.052*** (1.153)	-6.907*** (1.112)
Higher_Ed				-2.162* (1.184)	-2.017* (1.141)
Constant	15.20*** (1.411)	16.86*** (2.179)	20.56*** (2.422)	21.56*** (2.443)	16.65*** (2.326)
Sentiment	No	No	No	No	Yes
N	1000	1000	1000	1000	1000

Robust standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Notes: This table shows the effect of treatments on reported inflation expectations (the priors), relative to the *Flat* treatment. Regressions are estimated by OLS with robust standard errors. Data are winsorized at the 1% and 99% levels.

either incorporate incentives to enhance reliability and accuracy or correct for potential biases that may arise in their absence. However, two critical questions arise: What are we truly measuring when incentives are used, and do they allow us to accurately capture the genuine beliefs we seek to understand? While further research is needed to fully address these questions, our findings show that incentives change how households report their inflation expectations, resulting in more accurate expectations and disrupting the relationship between perceived and expected inflation (shown in Section 3.3.2).

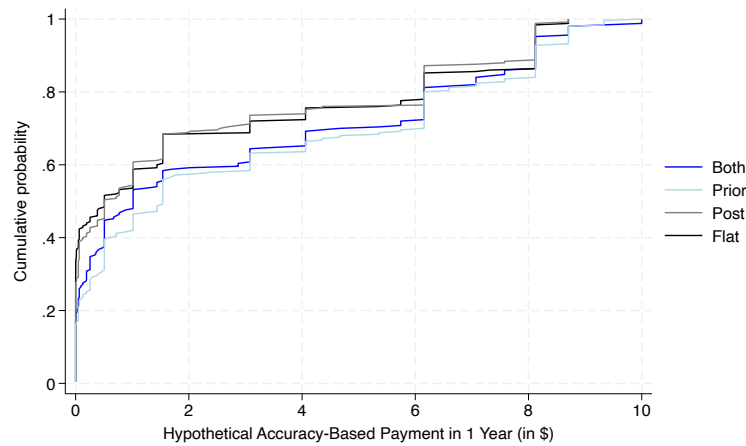
We compare data from each of our treatments to the most recent SCE microdata and the mean PCE forecast from the Survey of Professional Forecasters (SPF) in Table 3.4. While expectations under *Flat* (5.515%) closely track those from the SCE (6.178%), implementing marginal incentives in *Prior* leads to expectations (2.804%) that align more closely with professional forecasts from the SPF (2.11%).

We also consider how our various incentive schemes impact participants' hypothetical payoffs. To do this, we assume the Fed's forecast of median inflation for 2025 (π_{2025}) is correct in expectation. Using this as a basis for comparison, we calculate a participant i 's forecast error as $error_i = |\pi_{2025} - \mathbb{E}_i(\pi_{2025})|$ and her hypothetical bonus payment as $10 * (2^{-error_i})$. We depict the distribution of payoffs calculated this way across treatments in Figure 3.2 and explore the significance of these results in Table C.2.

Table 3.4. Comparing Experimental Data to SCE and SPF

Comparison	Group 1 Mean (Std. Dev.)	Group 2 Mean (Std. Dev.)	Difference	t-statistic	p-value
SCE vs. Flat	6.178 (12.942)	5.515 (24.910)	-0.663	-0.598	0.550
SCE vs. Prior	6.178 (12.942)	2.804 (14.313)	-3.374	-3.648	< .001
SPF vs. Flat	2.11 (0.286)	5.515 (24.910)	3.405	.796	0.427
SPF vs. Prior	2.11 (0.286)	2.804 (14.313)	0.694	0.966	0.778

Notes: This table compares data from SCE participants (first two rows) and from the Survey of Professional Forecasters (rows 3 and 4) to data from participants in *Flat* and *Prior* using two-sample t-tests. We restrict our focus to inexperienced SCE participants from the most recently-available microdata from September of 2023 when inflation was reported at 3.7%. For comparison, the most recent inflation report preceding our experiment was 2.9% (July inflation released August 14th). Data from the SPF are for the mean PCE inflation forecast for Q4 2024 to Q4 2025 (PCEB) from the Q3 2024 survey, which most closely aligns with our experimental time frame of September 2024 to September 2025. Note that the sample size for SPF (N=34) is considerably smaller than those of the SCE and our survey.

Figure 3.2. Hypothetical Payoffs for Point Forecasts of Inflation

Notes: The Figure shows how treatments impact the hypothetical payoffs of participants calculated comparing point forecasts formed before information provision the Fed's 2025 inflation forecast. This shows—assuming the Fed's forecast is correct in expectation—that expected payoffs are significantly higher for subjects facing marginal incentives. Data are winsorized at the 1% and 99% levels.

The punchline is that marginal incentives significantly increase hypothetical earnings. In *Both*, we predict in Table C.2 that payoffs will increase between approximately 24% ($p < .1$) in our baseline regression specification and 34% ($p < .05$) in a specification controlling for gender, education, and economic sentiment. In *Prior*, hypothetical earnings increase between 33% ($p < .01$) in our baseline specification and 49% in our full specification.

3.3.2 The Effect of Incentives on (In)Attention and Backward-Looking Heuristics

Why do incentivized expectations become more consistent with the SPF? A key factor appears to be (in)attention. In a similar spirit to Bracha and Tang (2024), who examine attention and perception errors in economic decision-making, we construct a measure

of inattention as an Absolute Perception Error (APE)—the absolute difference between a respondent’s perceived inflation and the most recent actual inflation rate available at the time of the survey.¹⁰ Intuitively, the further a respondent’s perception deviates from actual concurrent inflation, the less attention they are likely paying to inflation.

Column (1) of Table 3.5 shows that incentivized groups exhibit significantly lower inattention (or, equivalently, greater attention) to inflation. At the same time, incentives appear to weaken the relationship between perceived and expected inflation. Columns (2-5) of Table 3.5 present regression results where expected inflation is the dependent variable and perceived inflation is the independent variable. The findings indicate that in the incentivized group (*Prior*), this relationship is no longer statistically significant. Thus, inflation perceptions and expectations decouple with marginal incentives.

Table 3.5. Effects of Incentives on Inattention and Perception

	(1) All <i>APE</i>	(2) Flat $\mathbb{E}(\pi_{Prior})$	(3) Post $\mathbb{E}(\pi_{Prior})$	(4) Both $\mathbb{E}(\pi_{Prior})$	(5) Prior $\mathbb{E}(\pi_{Prior})$
Perception		0.337*** (0.079)	0.354*** (0.103)	0.270** (0.131)	0.165 (0.145)
Post	-2.034 (2.175)				
Both	-8.875*** (2.050)				
Prior	-11.479*** (1.914)				
Deflation	2.770** (1.456)	-29.908*** (2.815)	-23.802*** (2.668)	-19.134*** (2.223)	-14.827*** (1.754)
Male	-9.239*** (1.321)	-4.491** (2.258)	-2.081 (2.581)	-0.723 (1.836)	-2.170* (1.271)
Higher Ed	-3.589*** (1.362)	3.158 (2.288)	-3.996* (2.173)	1.596 (1.715)	-0.244 (1.529)
Constant	23.498*** (2.060)	9.496*** (2.272)	11.683*** (3.089)	6.225*** (1.578)	7.266*** (1.379)
<i>N</i>	1000	250	250	250	250

Robust standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Notes: This table shows the effect of treatment on Absolute Perception Error (APE) and the relationship between perceived inflation and prior point forecast. Regressions are estimated by OLS with robust standard errors. Data are winsorized at the 1% and 99% levels.

This result is particularly striking given that perceptions and expectations are elicited on the same survey page. One possible interpretation is that marginal incentives reduce reliance on simple backward-looking forecasting heuristics. Instead, incentivized respondents seem to

¹⁰The most recent PCE inflation preceding our experiment was 2.5% which was July inflation released on August 14th.

engage in more deliberate processing—and possibly acquisition—of information to generate more informed inflation forecasts.

As a result, we observe two simultaneous effects of incentives on inflation forecasts. First, incentives make forecasts more aligned with both realized inflation and professional forecasters, while also reducing disagreement among respondents. Second, incentives weaken the link between perceptions and expectations, suggesting that respondents move away from simple extrapolation of (perceived) past inflation. In other words, incentives shift the range of values respondents consider likely, leading them to form expectations within a more informed range.

Regarding the gender gap, we observe that women have noticeably higher APE (see Table C.3). This result complements Braitsch and Mitchell (2022), who construct a measure of inattention based on the consistency of responses to the SCE point and density forecast questions and show that women are less attentive than men when forming inflation expectations. However, we find that marginal incentives play a crucial role here as well, significantly reducing the APE gap.

3.3.3 Incentives and the Puzzle of the Gender Gap in Inflation Expectations

There is a long-standing strand of the survey-based belief literature that documents and attempts to rationalize gender differences in inflation expectations.¹¹ Concisely, female survey participants typically report significantly higher inflation expectations than do men. It is to this puzzle we now turn our attention. Our question is whether this puzzle survives the implementation of marginal incentives.

To do this, we estimate a series of OLS regressions for each treatment condition: *Flat*, *Post*, *Both*, and *Prior* where we project inflation expectations gathered before the information provision experiment (i.e. the priors) onto an indicator variable denoting whether a participant is female. This method enables us to independently assess the impact of gender within each specific treatment context.

The regression equation for each treatment T is specified as:

$$\mathbb{E}(\pi_{Prior,T}) = \beta_{0,T} + \beta_{1,T}\text{Female} + \epsilon_T$$

Additionally, we consider a final model where we pool our data across all four treatments and project inflation expectations onto a dummy variable capturing whether a participant was female, the incentive structure imposed on a participant, and the interaction of both terms. This specification allows us to test whether females exhibit a significantly different response to marginal incentives than males.

The interaction regression model is specified as:

$$\mathbb{E}(\pi_{Prior}) = \beta_0 + \beta_{1,i}\text{Treatment}_i + \beta_{2,i}\text{Female} + \beta_{3,i}(\text{Treatment}_i \times \text{Female}_i) + \epsilon_i$$

¹¹Note that we use the terms gender and sex interchangeably in this paper, as is common in related literature, though we recognize that they may not always align. For accuracy, the variable we use specifically measures sex.

where our coefficient of interest is β_3 .

The regression results, summarized in Table 3.6, reveal how incentives interact with gender to shape inflation expectations. In the absence of marginal incentives (*Flat*), female respondents have significantly higher inflation expectations (9.346, $p < .01$) than do their male counterparts. This finding is consistent with existing empirical literature that suggests that women often report higher inflation expectations or more pessimistic economic outlooks. This result also appears in *Post*, albeit muted and only marginally significant. Subjects know they would eventually face marginal incentives in *Post*, but are unsure of when exactly. It is possible that a spillover effect from the marginal incentives in the RCT portion of our survey heightened attention and effort sufficiently to mute the gender differences in inflation expectations to some extent.

Table 3.6. Effects of Incentives on the Gender Expectations Gap

	(1) Flat $\mathbb{E}(\pi_{Prior})$	(2) Post $\mathbb{E}(\pi_{Prior})$	(3) Both $\mathbb{E}(\pi_{Prior})$	(4) Prior $\mathbb{E}(\pi_{Prior})$	(5) All $\mathbb{E}(\pi_{Prior})$
Female	9.346*** (2.941)	4.779* (2.836)	2.681 (2.204)	1.082 (1.634)	
Post					3.970* (2.227)
Both					0.576 (1.999)
Prior					1.743 (1.802)
Female					9.346*** (2.941)
Post $\times$ Female					-4.568 (4.085)
Both $\times$ Female					-6.665* (3.675)
Prior $\times$ Female					-8.264** (3.364)
Constant	0.356 (1.489)	4.326*** (1.656)	0.932 (1.334)	2.099** (1.015)	0.356 (1.489)
N	250	250	250	250	1000

Robust standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Notes: This table shows the effect of treatments on reported inflation expectations (the priors) by gender. Regressions are estimated by OLS with robust standard errors. Data are winsorized at the 1% and 99% levels.

However, marginal incentives eliminate the significant difference in inflation expectations across genders. This is true for *Both* (2.68, $p > .1$) and *Prior* (1.08, $p > .1$). Further, we observe in column (5) that marginal incentives are killing the gender difference in expectations

observed in *Flat* because they act significantly more strongly on belief formation for females than they do for males (i.e., $Prior \times Female = -8.264$, $p < .05$, and $Both \times Female = -6.67$, $p < .1$).

These findings suggest that the puzzle of gendered expectations—where women report higher inflation expectations than men—diminishes under marginal incentives. Specifically, women appear to respond more strongly to incentives during belief elicitation, leading to more moderated and comparable expectations with men. This responsiveness effectively resolves the observed gender discrepancies in survey-based belief measures, as incentivized belief elicitation promotes more consistent and aligned inflation expectations across genders.

3.3.4 Effect of Incentives in Information Provision Experiments

We now focus on the role that marginal incentives play—or not—in a simple information provision experiment. We find that strategic incentives can effectively raise learning rates, suggesting that RCTs without marginal incentives may systematically underestimate the impact of information on beliefs.

After eliciting a point expectation of one-year-ahead inflation (alongside a few other measures), we provide each participant with a summary of the Fed’s outlook on how inflation might evolve in 2025. We then ask subjects to provide probabilistic forecasts for the evolution of food, gas, and aggregate prices. This creates distance between the information provision and the elicitation of posterior beliefs. We finally collect data from a binned inflation forecast to estimate a subject’s updated inflation expectation (“the posteriors”), which we treat as our measure of interest throughout this section.¹² Recall that we adopt the same point and bin elicitation strategies followed by the NY Fed in its Survey of Consumer Expectations.¹³

As noted by Haaland et al. (2023), information provision experiments measuring beliefs and belief updating via quantitative measures typically quantify the extent to which respondents adjust their beliefs toward new information conveyed via some signal. They call this the *learning rate*, which one can estimate using

$$Updating_i = \beta_0 + \beta_1(Treatment_i \times PercGap_i) + \beta_2Treatment_i + \beta_3PercGap_i + \epsilon_i \quad (3.2)$$

where $Updating_i$ is the distance between respondent i ’s posterior and prior one-year-ahead inflation expectation, $Treatment_i$ is an indicator variable denoting which incentive structure a respondent faced, and $PercGap_i$ (Perception Gap) is the distance between the Fed’s forecast of median PCE inflation in 2025 and respondent i ’s prior for the same. Given these

¹²We depict these expectations in Figure C.1 and explore whether marginal incentives impact the distributions of these expectations collected across treatment in Table C.4. For the sake of brevity, we report both in Appendix C.1. Similar to what D’Acunto et al. (2023) demonstrates for the SCE data, we find that the binned inflation forecasts exhibit less disagreement and a lower mean expected inflation than point forecasts.

¹³There is evidence from Becker et al. (2023) that the number, center, and width of bins can significantly alter expectations provided by survey respondents. Though beyond the scope of this paper, an interesting question for future research is whether and how this might interplay with marginal incentives.

definitions, comparing across β_1 captures the extent to which incentive structure drives belief updating relative to our baseline treatment *Flat*, β_2 captures the average treatment effect on respondents' beliefs that does not depend on individual priors, and β_3 measures the extent to which changes in beliefs depend on the perception gap.¹⁴

Table 3.7. Effect of Incentives on Learning Rates

	(1) Updating	(2) Updating	(3) Updating	(4) Updating
$\beta_{1,Post}$	0.0344*** (0.0113)	0.0345*** (0.0113)	0.0348*** (0.0112)	0.0326*** (0.0111)
$\beta_{1,Both}$	0.0262* (0.0135)	0.0265** (0.0134)	0.0269** (0.0134)	0.0261** (0.0126)
$\beta_{1,Prior}$	0.0189 (0.0230)	0.0190 (0.0231)	0.0194 (0.0232)	0.0195 (0.0226)
$\beta_{2,Post}$	-0.448* (0.244)	-0.444* (0.244)	-0.452* (0.243)	-0.492** (0.237)
$\beta_{2,Both}$	-0.212 (0.240)	-0.232 (0.242)	-0.239 (0.242)	-0.191 (0.237)
$\beta_{2,Prior}$	-0.0648 (0.239)	-0.0881 (0.241)	-0.0904 (0.241)	-0.0511 (0.236)
β_3	0.910**** (0.00772)	0.909**** (0.00777)	0.909**** (0.00779)	0.913**** (0.00775)
Male		-0.155 (0.166)	-0.157 (0.166)	-0.0649 (0.167)
Higher_Ed			0.0598 (0.166)	0.0335 (0.163)
Constant	0.0185 (0.177)	0.102 (0.200)	0.0756 (0.218)	0.485** (0.246)
Sentiment	No	No	No	Yes
N	1000	1000	1000	1000

Robust standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Notes: This table shows the effect of treatments on learning rates. These are relative to our baseline treatment *Flat*. Regressions are estimated by OLS with robust standard errors. Data are winsorized at the 1% and 99% levels.

Table 3.7 shows results of estimating several versions of Equation 3.2, where $\beta_{1,i}$ is primarily our coefficient of interest. Column (1) shows our baseline specification with no additional controls, and columns (2) through (4) include additional controls for sex, educational attainment, and economic sentiment. Regardless of specification, $\beta_{1,Post}$ is significant at the

¹⁴Haaland et al. (2023) argue that if priors are balanced across treatment, the researcher could use the posterior as the dependent variable. We cannot do that here, since treatment variation can induce systematic differences in the prior.

1% level, indicating that imposing marginal incentives in our simple information provision experiment led to significantly more belief updating. The same effect is true for *Both*, where the accuracy of both prior and posterior beliefs about inflation is incentivized. It is therefore not too surprising that the learning effect (captured by $\beta_{1,Both}$) is roughly 30% smaller across all specifications and significant at only the 10% (baseline specification, column (1)) or 5% (columns (2) - (4)) levels.

Our analysis demonstrates that implementing marginal incentives significantly enhances belief updating among subjects. Specifically, the positive and statistically significant coefficient for the interaction term ($\beta_{1,Post}, \beta_{1,Both}$) indicates that incentives designed to promote forecast accuracy significantly amplify the magnitude of belief updating in response to discrepancies between their prior beliefs and the Federal Reserve’s forecasts. This suggests that strategic incentives can effectively bridge perception gaps, leading to belief formation that is more responsive and consistent with professional forecasts. More importantly, this suggests that RCTs implemented without marginal incentives may systematically underestimate the impact of information on beliefs, especially in environments where expectations are well-anchored, such as during periods of low inflation (Weber et al., 2025; Bracha and Tang, 2024). Implementing marginal incentives suggests a potentially stronger ability of central bank forecasts to coordinate and guide inflation expectations.

We also consider how marginal incentives changed the distribution of probabilities participants assigned to each possible inflation bin. Recall, values assigned to each bin when forming a bin forecast denote the participant’s belief about the likelihood of realized inflation falling into that particular bin. Figure 3.3 depicts the average weight assigned to each of the ten possible inflation bins presented to subjects who faced marginal incentives (blue dashed line) or did not (red solid line). Interestingly, imposing marginal incentives shifts significantly more weight to the bin containing the Fed’s median forecast for 2025 inflation communicated in the information intervention and significantly reduces weights assigned to more extreme bins. This further reinforces our result that marginal incentives significantly alter the efficacy of information provision in such experiments.

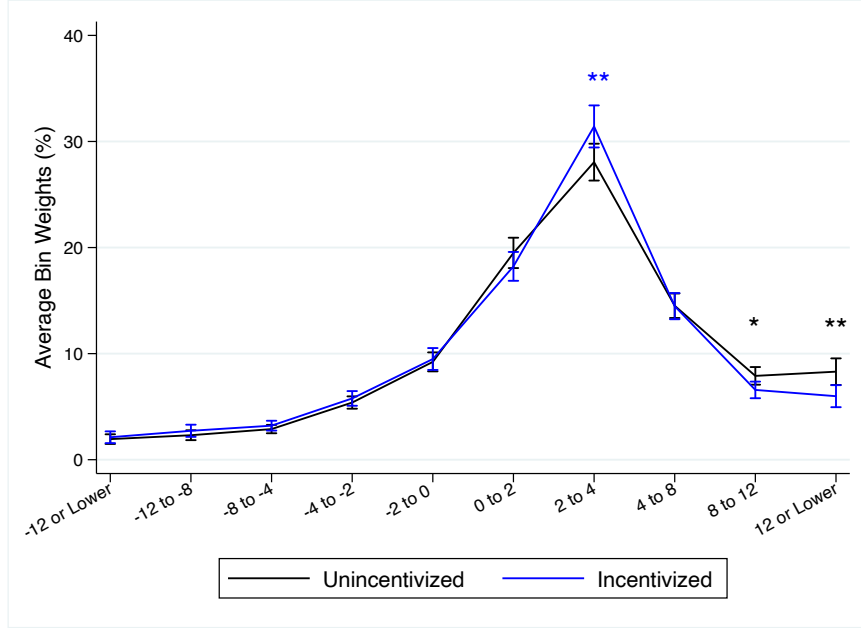
3.3.5 Incentives and Effort

An essential consideration in survey-based research is the amount of effort participants invest when responding to questions, particularly when eliciting complex beliefs such as inflation expectations. Or similarly, how compliant participants are in considering provisioned information in RCTs (Knotek et al., 2024).

We quantify effort using survey *completion time*, a common metric in survey research used to approximate the cognitive resources participants allocate to answering questions (Malhotra, 2008). We design the survey to take approximately five minutes, but we anticipate variation based on individual differences in reading speed, comprehension, and the effort invested in considering responses. By comparing completion times across different incentive treatments, we can assess whether marginal incentives motivate participants to devote more time—and thus presumably more cognitive effort—to the survey tasks.

We estimate a series of ordinary least squares (OLS) regressions with data winsorized at the 5% and 95% levels to analyze the impact of marginal incentives on completion time. The regression equation is specified as

Figure 3.3. Average Bin Weights across Incentives



Notes: The Figure shows the average weight participants placed into the respective bins, distinguishing between unincentivized (*Flat* and *Prior*) and incentivized (*Post* and *Both*) treatments. Data are winsorized at the 1% and 99% levels.

$$\text{CompletionTime}_i = \alpha + \gamma_1 \text{Post}_i + \gamma_2 \text{Both}_i + \gamma_3 \text{Prior}_i + \beta \mathbf{X}_i + \epsilon_i, \quad (3.3)$$

where CompletionTime_i is the total time (in seconds) participant i takes to complete the survey. Post_i , Both_i , and Prior_i are dummy variables indicating the incentivized treatment group to which participant i is assigned, with the *Flat* treatment serving as the reference group. $\mathbf{X}_i$ is a vector of control variables, including participant gender (*Male*), education level (*Higher_Ed*), and forward-looking economic sentiment (included in column (4)).¹⁵

We report regression results in Table 3.8.

Coefficients for *Both* and *Prior* are positive and statistically significant across all specifications, indicating that participants in these groups take significantly longer to complete the survey compared to those in the *Flat* treatment. Specifically, participants in the *Both* treatment spend approximately 104 to 111 seconds more on the survey than those in the *Flat* group—a substantial increase given the survey’s average completion time. Those in the *Prior* treatment take about 50 to 58 seconds longer than participants in the *Flat* treatment. Although the coefficients for the *Post* treatment are positive, they are not statistically significant, suggesting that marginal incentives applied only after the information provision do not significantly affect overall completion time.

These results support the hypothesis that marginal incentives enhance participant effort during belief elicitation, particularly when the incentives are applied at the initial stages of the survey, as in the *Prior* and *Both* treatments. The increased completion times indicate

¹⁵We show the same results without winsorizing in Table C.5, in Appendix C.1.

Table 3.8. Effect of Incentives on Effort

	(1) Completion Time	(2) Completion Time	(3) Completion Time	(4) Completion Time
Post	19.94 (25.86)	23.05 (25.71)	27.18 (25.64)	29.24 (25.68)
Both	110.8**** (26.80)	104.6**** (26.72)	106.4**** (26.53)	104.0**** (26.50)
Prior	58.43** (25.29)	51.37** (25.13)	50.11** (25.16)	53.37** (25.20)
Male		-70.62**** (18.63)	-67.44**** (18.60)	-72.43**** (18.56)
Higher_Ed			-56.19*** (18.26)	-59.62*** (18.26)
Constant	567.4**** (18.39)	599.0**** (19.88)	625.0**** (21.97)	585.9**** (25.59)
Sentiment	No	No	No	Yes
N	1000	1000	1000	1000

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Notes: This table shows the effect of treatments on effort, as proxied by completion times. Regressions are estimated by OLS with robust standard errors. Data are winsorized at the 5% and 95% levels.

that participants are investing more effort into responses, leading to more thoughtful belief formation.

The regression results also reveal significant effects of participant gender and education level on completion time. The coefficient for male participants is negative and highly significant across specifications, indicating that males spend approximately 67 to 72 seconds less on the survey than females. This suggests that female participants generally invest more time and effort into completing the survey tasks. Additionally, participants with at least an undergraduate degree spend about 56 to 60 seconds less on the survey compared to those without higher education. This may reflect greater familiarity with the survey content or more efficient processing of the information among more educated participants. The inclusion of economic sentiment in column (4) does not substantially alter the coefficients of interest, and the main findings regarding the impact of marginal incentives on completion time remain robust.

3.4 Discussion

Our experiment demonstrates that marginal incentives significantly impact the elicitation of macroeconomic beliefs and learning rates in information provision experiments. Specifically, imposing incentives leads to household inflation expectations that are more consistent with expectations of experts, with respondents predicting less extreme values and exhibiting

less cross-sectional forecast disagreement. Furthermore, these incentives increase the rate at which participants update their beliefs in response to new information, suggesting that central bank communication can more effectively coordinate expectations in environments where households have higher incentives to pay attention to inflation. Our findings suggest that central banks can boost the effectiveness of their communication policies by exploring innovative ways to incentivize households. Options include decreasing the cost of paying attention, for example, by improving their communication strategy, or alternatively, by raising the (perceived) benefits of information through improving financial literacy.

These findings challenge the current reliance on unincentivized or flat-fee structures in most survey-based macroeconomic research, highlighting potential biases in elicited beliefs when incentives are absent. The substantial reduction in potential forecast errors and heightened learning rates observed with marginal incentives indicate that incentivized elicitation provides a more reliable measure of household expectations, which are critical for policymaking.

Notably, the same incentive structures that lead to higher learning rates also close the gender gap in inflation beliefs in point forecasts of one-year-ahead inflation, offering a simple resolution to a long-standing puzzle in the belief-based macroeconomic literature. This suggests that marginal incentives can mitigate systematic biases observed in survey-based beliefs elicited via flat-fee incentives.

Incorporating marginal incentives into survey design leads to more consistent measurements of inflation expectations across respondents and brings household inflation forecasts closer to professional forecasters' consensus views. This reduction in cross-sectional disagreement and convergence toward expert predictions suggests that incentivized surveys may provide policymakers with more reliable measurements for understanding household expectations—a crucial input for those focused on managing inflation dynamics and developing economic forecasting models. Notably, this change in incentive structure need not increase the cost of collecting expectations. Further, policymakers and researchers can leverage our approach to quantify the extent of measurement error due to the absence of incentives.

Our findings also relate to the literature on the impact of external conditions on RCTs. Studies suggest that extraneous information conditions can mute the effects observed in RCTs. For instance, during periods when attention to inflation and monetary policy is high—as has recently been the case—the efficacy of information provision may appear diminished in unincentivized settings. Our results from the flat-fee setting align with this, showing that information provision without incentives does not significantly affect beliefs. However, information provision becomes more effective in shifting expectations under marginal incentives. This highlights the importance of incentive mechanisms in accurately assessing the impact of policy communications.

While our study underscores the benefits of incorporating marginal incentives, it is important to consider potential drawbacks. One concern is that complex incentive schemes may induce behavior that does not reflect genuine beliefs but rather strategic reporting, such as looking up information online. Participants might engage in actions aimed at maximizing their payoffs rather than truthfully revealing their expectations.

For example, Grewenig et al. (2022) find that providing incentives does not impact beliefs

about personal earnings—which are readily available to participants—but improves beliefs about average public school spending, a less accessible piece of information for the average respondent. The authors highlight a trade-off between increased respondent effort and the risk of inducing online search activity when incentivizing beliefs in online surveys. While our experimental setup minimizes this risk by controlling the information environment, it remains a potential issue in other survey contexts.

Additionally, Danz et al. (2022) provide evidence that complex incentive schemes can lead to misunderstandings, potentially resulting in less truthful reporting. They find that truthful reporting increases when information about incentives is absent compared to a baseline condition that provides full details about how incentives are determined using a binarized scoring rule (BSR). This suggests that overly complicated incentive mechanisms may confuse participants, undermining the very accuracy they are intended to enhance.

Therefore, while marginal incentives can improve data quality by motivating participants to invest more effort and report more consistent beliefs, careful consideration must be given to the design of these incentives. Simplicity and transparency are crucial to avoid inducing strategic behavior or confusion that could compromise the integrity of the data.

In conclusion, our study suggests that incorporating marginal incentives into surveys positively enhances elicited beliefs, which could improve the robustness of empirical findings in macroeconomic research. By motivating participants to engage more deeply when forming beliefs, incentivized mechanisms can lead to more reliable data. Future studies should consider integrating such mechanisms to improve data quality, while also being mindful of the potential pitfalls associated with incentive complexity. Balancing the benefits of increased effort and accuracy against the risks of strategic behavior and misunderstanding is essential for advancing survey-based measures and of economic expectations and informing more effective policy decisions.

Appendix to Chapter 1: The Messenger Matters

A.1 Details on Twitter Data and Additional Evidence

A.1.1 Cleaning Twitter Data

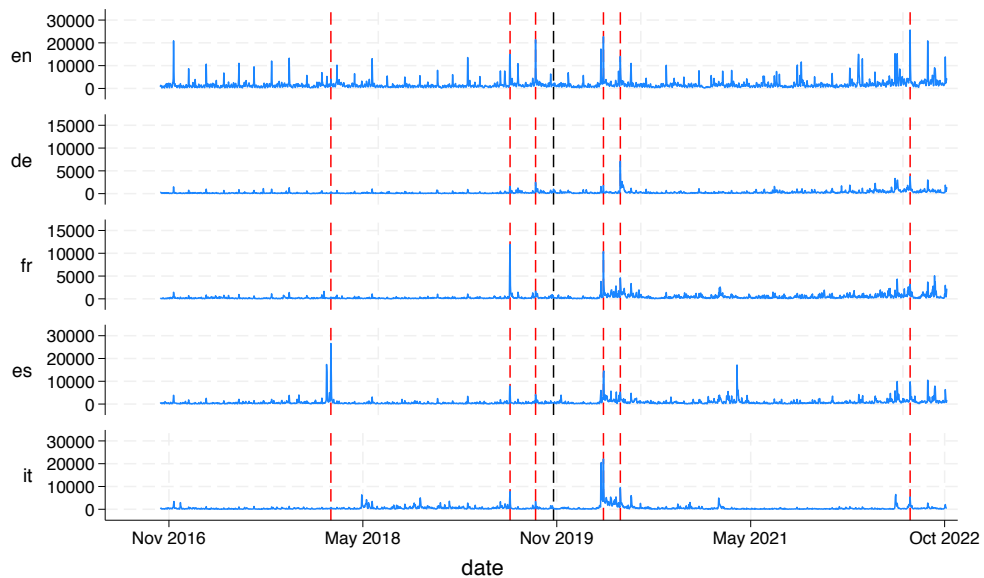
Tweets are collected between November 2022 and March 2023. Before any data cleaning, the dataset consists of just over 11 million tweets. First, I drop tweets that are unrelated to central banking. I start by manually identifying keywords that indicate unrelated tweets, such as any tweets about the English Cricket Board, which also abbreviates to ECB. This cleaning step is supported by two different word clouds. One visualizes the most frequent words for retained tweets, and may show any key words that are likely from unrelated tweets (e.g., “cricket”, or names of cricket players), and the other visualizes most frequent words of dropped tweets, to inspire further keywords based on tweets are dropped for my final sample. The final word cloud for all languages of the cleaned sample is shown in Figure 1.1 and Figure A.2. These word clouds indicate that the final sample indeed contains at a decisive majority tweets about central banking. This leaves just over 8 million (8,031,937) tweets, which I clean further, and translate to English using Google Translate’s Neural Machine Translation. Specifically, I lower cases in each tweet and remove any special signs, emojis, digits, links to websites, in-tweet mentioned other users (indicated by preceding “@”) except for “lagarde” and “ecb”, and fully spell out contracted words (e.g., “can’t” to “cannot” or “we’ll” to “we will”). I translate all tweets, including English tweets, as this helps eradicate spelling mistakes, and ensures consistent, American English spelling of words, which is important for dictionary-based natural language processing (NLP) approaches.

A.1.2 Descriptive Analysis of Twitter Data

Tweet volume The final sample contains just over 8 million tweets, whereby daily tweet volumes vary over time and by language (see Figure A.1). The sample of English

tweets stands out by its volume making up almost half of the entire sample (46%). This is not surprising as English is the official language of the ECB, and the lingua franca for policymakers and economists.

Figure A.1. Daily Tweet Volume by Language



Notes: Daily volume of tweets (using CET as this is the time zone corresponding to the headquarter of the ECB). The dashed grey vertical line indicates the changeover in presidency (November 1, 2019). Dashed red vertical lines indicate the following events (from left to right): Eurogroup's support for de Guindos (February 19, 2018), announcement of Lagarde as incoming ECB president (July 3, 2019), the day after PEPP announcement (March 19, 2020), German constitutional court ruling (May 5, 2020), ECB raising rates for the first time in 11 years (July 21, 2022).

Idiosyncrasies of Language Samples and the Validity of the Proxy for Nationality

The daily tweet volume maxima (i.e., peaks) of tweet samples split by language coincide strongly with events of severe national interest. This is strong evidence in favor of the validity of proxying nationality with tweet language. Peaks in the time series across all samples generally refer to a decision by the ECB or an announcement of new ECB board members. Interestingly, the peaks of the time series are different for each language, and coincide with matters that have particular importance to different countries. For instance, the announcement of Lagarde as the incoming president of the ECB caused the peak in French data in July 2019, while the Eurogroup giving support to the candidacy of the Spanish vice president, de Guindos, in February 2018 caused the peak in the Spanish data. The peak in the German data coincides with German constitutional judges ruling that the ECB's asset purchasing programme partly violated the German constitution in May 2020. The peak in Italian data occurred in March 2020, one day after the announcement of the PEPP (Pandemic emergency purchase programme), which caused Italy's borrowing costs to fall sharply after the early and hard onset of Covid-19 in Italy. The peak in English data highlights how English tweets might be different from any other languages by marking the day on which the ECB raised interest rates for first time in 11 years (July 2022). This suggests that English is used relatively more for debating technicalities of EA-wide monetary policy

than other languages that show a stronger focus on news concerning specific countries. At the same time, this means that analyzing non-English tweets seems important to understand how particularly non-experts respond to central bank communication. Table A.1 provides an overview of all peaks.

Table A.1. Peaks in Tweet Volume by Language

Language	Peak Volume	Peak Date	Event
English	25,624	21 Jul 2022	ECB raising rates for first time in 11 years
German	7,098	05 May 2020	German constitutional court ruling
Spanish	26,599	19 Feb 2018	Eurogroup's support for Luis de Guindos
French	11,905	03 Jul 2019	Announcement of Lagarde as incoming ECB president
Italian	22,050	19 Mar 2020	Day after PEPP announcement

Notes: Tweet volume is defined as daily volume of tweets (using CET as this is the time zone corresponding to headquarter of the ECB).

Sentiment measure Sentiment of tweet content is measured by a dictionary-based method that provides a continuous estimate of how positive or negative a text snippet is. A sentiment score is computed for each tweet using the polarity measure of the Python package TextBlob, which is based on the Princeton University's WordNet lexicon (Loria, 2018). Sentiment scores are estimated on tweets that are cleaned and translated (and thus spell-checked). This ensures that the same (English) dictionary can be used to create the sentiment scores. I normalize this sentiment indicator to assign values between 0 (most negative) and 1 (most positive) to each tweet, where neutral text gets a value of 0.5. The original polarity sentiment indicator returns a value between -1 and 1, where the most positive (negative) text obtains a value of 1 (-1) and neutral text gets a value of 0.

Sentiment Across Languages To be able to make meaningful comparisons of beliefs across languages, it is important that no language shows idiosyncratic differences in the measure of tweet sentiment. Given the strong similarity in the distribution of tweet sentiment, this seems to be indeed the case. An overview of tweet sentiment across languages can be found in Table A.2. The average and standard deviation of tweet sentiment are remarkable similar across languages.¹

Time Trends The sample covers six years; three entire years for each president. Many things happened in these six years, including digital advancements and increased computerization, all accelerated by the Covid-19 pandemic that moved much of communication online. Thus, it is not surprising that the number of tweets and the number of Twitter users has increased between the two three-year periods. The tweet volume increased by 53.5% between the three years of Draghi's presidency and the three years of Lagarde's. The total number of Twitter users increased even more: by 74.9%. Table A.3 gives an overview of how the number of tweets and the number of Twitter users have changed between the two

¹Average tweet sentiment is close to neutral between 0.52 and 0.53, and the standard deviation is either 0.10 or 0.11. Even the 10th and the 90th percentiles of tweet sentiment do not vary considerably across countries, with values for the 10th percentile ranging between 0.42 and 0.45, and for the 90th between 0.63 and 0.65, respectively. It is noteworthy, however, that across languages most tweets have rather neutral sentiment around 0.5, and that tweets are slightly more positive than negative.

Table A.2. Sentiment of Tweets by Language

Language	N	Mean	Std. dev.	p10	p90
English	3,697,934	0.53	0.10	0.45	0.65
German	565,854	0.52	0.10	0.42	0.63
Spanish	1,729,280	0.53	0.10	0.44	0.65
French	823,712	0.53	0.10	0.44	0.64
Italian	1,215,157	0.52	0.11	0.42	0.65

Notes: Tweet Sentiment ranges from 0 to 1, where more positive (negative) tweets are closer to 1 (0) and neutral tweets score 0.5. Means and standard deviations are based on all available individual tweets by language between 2016–2022.

three-year periods by language. Blue font highlights when a nationality is in the ingroup with the ECB president. For the French sample, the increase in the number of tweets was by far the largest (145.3%). This coincides with their transition into the ingroup. Similarly, the number of French accounts grew substantially between the three years of Draghi’s presidency and Lagarde’s (by almost 93%). Only the number of German accounts grew more (by 127%), but from the lowest initial level. And even though the number of distinct users in the French sample is still below average (even when excluding English users), and the ranking of languages by the number of distinct users did not change between the two periods of presidency, it is noteworthy that the Italian sample—the Twitter users that fell out of the ingroup—grew the least (by only 30%). Similarly, Italian tweets saw the smallest growth (only 23.6%). Of course, there is a chance that this is a coincidence or merely reflecting heterogeneous levels of digitization across countries that have nothing to do with the ECB presidency, but this could also be another indicator of the ingroup behaving differently than the outgroup.

Table A.3. Number of Tweets and Distinct Twitter Users

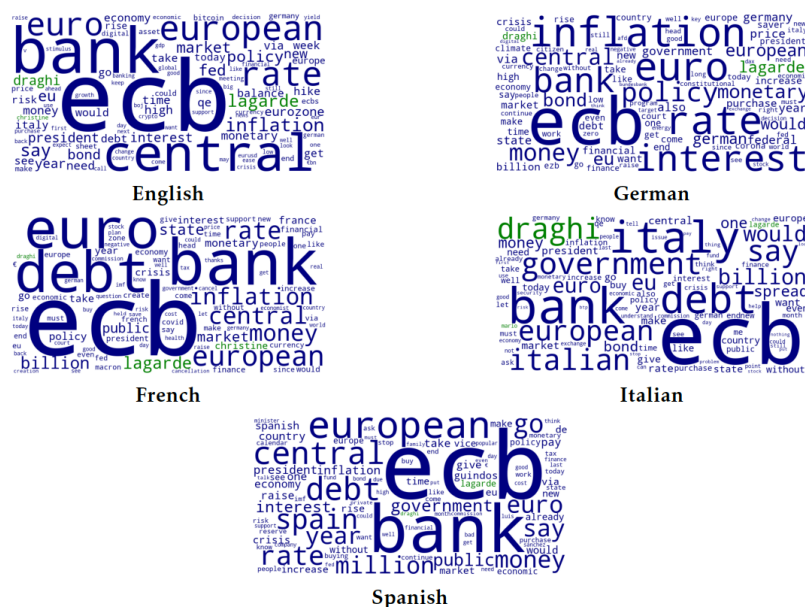
	Presidency		Tweet Growth
	Draghi	Lagarde	(User Growth)
<i>English</i>	1,595,810 (279,482)	2,102,124 (499,940)	31.7% (78.9%)
<i>Spanish</i>	623,046 (153,419)	1,106,234 (257,135)	77.6% (67.6%)
<i>French</i>	238,455 (62,737)	585,257 (120,942)	145.4% (92.8%)
<i>Italian</i>	543,504 (60,403)	671,653 (78,873)	23.6% (30.6%)
<i>German</i>	168,006 (29,903)	397,848 (67,829)	136.8% (126.8%)

Notes: The Table reports number of tweets and distinct Twitter users (in brackets) by language and presidency, where “Draghi” refers to tweets posted during 2016–2019 until the switchover in presidency in November 2019, and “Lagarde” refers to tweets posted afterwards (2019–2022). Languages are ranked by the number of distinct Twitter users (ranking remains the same across presidencies). The blue font indicates ingroup periods.

A.1.3 Additional Tables and Figures

The Sample The word clouds shown in Figure A.2 indicate that the cleaning of tweets is successful in ensuring that tweets are indeed about central banking. In addition, the figure visualizes how different languages vary in their focus of discussions on Twitter, as discussed in the main part of the paper.

Figure A.2. Word Clouds by Language



Notes: The figures show the 100 most frequent words between 2016-2022. Samples are split by original tweet language. Word size indicates word frequency. The following terms are highlighted in green for better visibility: “christine”, “lagarde”, “mario” and “draghi”. Word cloud is based on cleaned, translated and lemmatized tweets. Stopwords are removed. No ngrams are included.

Table A.4. Ingroup Effect on Information Supply on Twitter

	(1) All	(2) Experts	(3) Non-Experts
Ingroup=1	0.105*** (0.0179)	0.0244** (0.0103)	0.0778*** (0.0190)
ES	0.256*** (0.0179)	0.129*** (0.0103)	0.243*** (0.0190)
FR	-0.00138 (0.0200)	-0.0944*** (0.0115)	0.0195 (0.0212)
IT	0.0887*** (0.0200)	0.0923*** (0.0115)	0.0903*** (0.0212)
Constant	0.138*** (0.0127)	0.212*** (0.00725)	0.142*** (0.0134)
N	200	200	200
R-squared	0.582	0.738	0.505

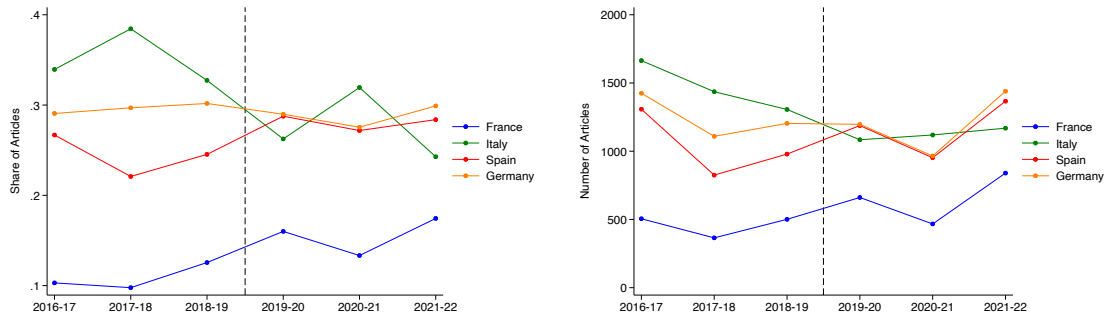
Notes: The table shows OLS regression results of being in the ingroup with the ECB president on the share of tweets by language per 6-week press conference (PC) cycles, controlling for language. German acts as the baseline language. The number of observations reflects the 4 languages and 49 press conferences in the cycle, where one press conference cycle is split in partly being under Draghi's presidency and partly under Lagarde's, making it a total of 50 president-PC cycles combinations. The (non-)expert classification follows the benchmark in Ehrmann and Wabitsch (2022). Standard errors in parentheses. Significance level is indicated by stars: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A.1.4 Does Print Media Confirm Increased Information Availability?

Twitter covers information availability beyond mere newspaper coverage, as it better reflects how exposed individuals are on average to given news story. However, to preclude doubts that information availability via print media provides a different picture to the results from Twitter, I analyze a sample of newspaper articles sourced from LexisNexis. The selection criteria mirror the Twitter sample: Articles are included in the sample if they contain the keywords "ECB", "European Central Bank" or the translated equivalents in the respective languages. For each nationality, the source of newspaper articles must be in the respective country and the language of the article must be the local language. The sample is restricted to only newspaper articles and the sample spans from November 1, 2016 to October 31, 2022, which reflects exactly three years of presidency for both Mario Draghi and Christine Lagarde. To avoid sample fluctuations over time that are due to LexisNexis-related data availability rather than reflecting actual changes in newspaper articles, I limit my sample to a few major newspapers for each country that are consistently available for the entirety of the sample, while ensuring political perspectives of selected newspapers are balanced. For Italian news, I include "Corriere della Sera" and "La Nazione". The French sample is comprised of "Le Figaro", "La Tribune", and "La Croix". The German sample contains "BILD" (incl. Sunday and regional editions), "Süddeutsche Zeitung" (incl. regional editions), and "Der Tagesspiegel". Finally, the included Spanish newspapers are "El Mundo" and "El País".

To assess newspaper article volume over time, the switchover in presidency acts as a pseudo-treatment, allowing to analyze the impact of the policymaker's nationality on national news with a difference-in-difference style analysis. While newspaper volume for Germany and

Figure A.3. National Newspaper Articles Over Time



Notes: The left panel plots countries' share of articles each year, while the right panel shows the volume of national newspaper articles over time. Each point is based on the volume of newspaper articles over 12 months (from November 1 until October 31 of the following year). The vertical grey dashed line indicates the switch in presidents.

Spain does not change much between the presidency of Draghi compared to Lagarde, Italian news volume declines during Lagarde's term time, while it increases for France. Figure A.3 visualizes these trends. Table A.5 shows the estimated effect, controlling for time- and country-specific effects. The ECB president's nationality significantly increases the share of their corresponding national newspaper volume by 6.1 percentage points (or by roughly 272 articles).

In sum, also with traditional news, I find that the ECB president's nationality raises traditional national newspapers reporting. This confirms that there are positive ingroup effects on the availability of information, in line with the increased information availability seen on Twitter.

Table A.5. Effect of ECB President's Nationality on Newspaper Volume

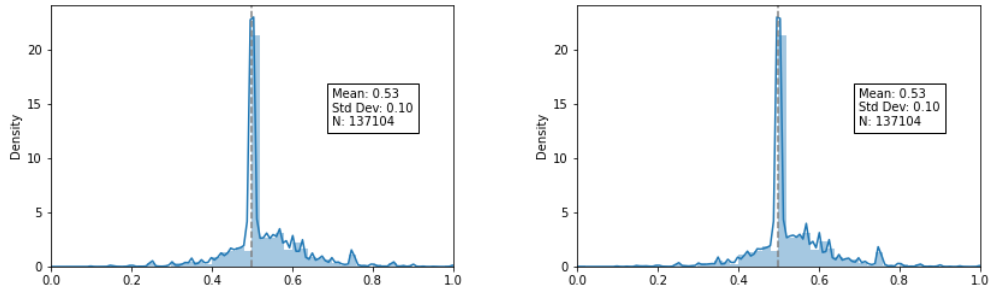
	(1) Share Of Articles	(2) Share Of Articles	(3) Number of Articles	(4) Number of Articles
Ingroup	0.061*** (0.018)	0.061** (0.021)	271.8*** (68.1)	271.8** (71.6)
2017-2018	-0.000 (0.019)	0.020 (0.028)	-291.8*** (78.2)	-184.0 (97.3)
2018-2019	-0.000 (0.012)	0.005 (0.025)	-228.0*** (67.2)	-181.0 (155.4)
2019-2020	-0.000 (0.011)	-0.010 (0.011)	-193.0*** (53.8)	-212.0* (94.3)
2020-2021	-0.000 (0.017)	0.005 (0.041)	-349.8*** (66.0)	-291.5 (144.3)
2021-2022	0.000 (0.015)	-0.013 (0.031)	-21.5 (73.2)	-80.0 (124.0)
ES	-0.030** (0.013)		-120.2* (56.8)	
FR	-0.191*** (0.013)		-802.8*** (67.0)	
IT	-0.010 (0.020)	0.180*** (0.021)	-62.9 (76.6)	739.8*** (71.6)
Constant	0.292*** (0.007)	0.100*** (0.013)	1404.0*** (42.2)	578.7*** (104.7)
Observations	24	12	24	12
R ²	0.925	0.952	0.941	0.970

Notes: Column (1) shows the effect of being in the ingroup with the ECB president on the share of newspaper articles, controlling for time- and country-FE. Column (2) shows the same but limiting the sample to only nations that switch between the in- and outgroup (France and Italy). Similarly, columns (3) and (4) show the same for actual number of articles instead of shares. All regressions are estimated by OLS, and robust standard errors are shown in brackets. *Ingroup* is a dummy variable that equals 1 for the country of origin of the concurrent ECB president. Omitted baseline category for ingroup is the outgroup, for time periods it is 2016-2017, for countries it is Germany or France. Stars correspond to the following p-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.1.5 Empirical Estimation of Belief Updating

Priors and Posteriors To assess individual belief updating, I compare the sentiment of tweets before and after ECB press conferences. For users who tweet both before and after a conference, the posterior is defined as the sentiment of the first post-conference tweet: $Posterior_i = Sentiment_{i,t}$, where i represents an individual Twitter user and t is the time of the tweet. The prior is the sentiment of the user's last tweet during the ECB's quiet period (7 days before the conference) just before the press conference: $Prior_i = Sentiment_{i,t-1}$. This approach captures responses to the press conference and avoids picking up other ECB communications due to the quiet period's restricted communication. Figure A.4 shows the distribution of priors and posteriors.

Figure A.4. Distributions of Priors and Posteriors



Notes: The plots show the distribution of variables (priors and posteriors) by combining a histogram of each variable's values with a kernel density estimation (KDE) plot. The dotted vertical grey line goes through 0.5, which indicates neutral values across variables. Any values to the right (left) of the dotted line can be interpreted as positive (negative). The text boxes show the mean and standard deviation of the variables, as well as the number of observations. The left (right) panel shows normalized values of priors (posteriors) in the final sample of disaggregated analyses.

Empirical Strategy To empirically test the belief updating using Twitter data, I use an approach similar to Coibion et al. (2021). While the authors focus on the impact of the prior on the posterior in a single period, I analyze multiple updates of individuals over time, where t represents the time period corresponding to a specific press conference. Therefore, instead of including a constant, I include fixed effects for each updating event (i.e., press conferences). These fixed effects swallow any impacts the actual signals have on the updated posterior and simultaneously avoid making any assumptions about the signals.

$$Posterior_{i,t} = \beta_1 Prior_{i,t} + \delta' D_t + \epsilon_{i,t} \quad (A.1)$$

Coibion et al. (2021) assume that the distance between the prior's coefficient to 1 is explained by the effect of the missing signal. I first use this assumption to understand updating at different points in time first, and later incorporate actual signal values of press conferences for the main specifications. With the assumption, the prior coefficient is interpreted as follows: If the coefficient β_1 is equal to 0, the signals are interpreted as fully informative and the agents neglect their priors. Similarly, if $\beta_1 = 1$ the signals are fully uninformative,

and if $0 < \beta_1 < 1$ the signals are partially informative. Thus, the larger the prior coefficient β_1 (i.e., closer to 1), the less agents respond to novel information from signals, and the more individuals rely on their prior beliefs when forming posterior beliefs. Note that a rational, Bayesian agent incorporates both the prior and the signal when forming their posterior belief, in proportion to their respective precision. Prior precision refers to the confidence or uncertainty assigned to the agent's prior belief, and signal precision is determined by signal accuracy. Since this information on priors and signals is not available in the observational data, no conclusions on whether updates are Bayesian or not can be made here.

To estimate ingroup effects, I include the interaction with a dummy ($Ingroup_{i,t}$) that indicates whether an individual i is in the ingroup. This shows whether and how individuals of the ingroup update differently than outgroup individuals. A negative coefficient of the interaction term suggests the ingroup relies less on their priors and more on new information more into their beliefs.

$$Posterior_{i,t} = \beta_1 Prior_{i,t} + \beta_2 Prior_{i,t} * Ingroup_{i,t} + \beta_3 Ingroup_{i,t} + \delta' D_t + \epsilon_{i,t} \quad (A.2)$$

Another advantage of avoiding the quantification of signals is that it allows me to flexibly define different updating horizons. Specifically, I first use Equation (A.2) to assess how individuals update during quiet periods, which refer to the week before each press conference, where the ECB purposefully limits its communication with the public. Therefore, the expected coefficients for the prior should be higher than in the previous specification. This is because if there is less new information about central banking disclosed, beliefs should resemble their priors more. Second, I assess how individuals update outside quiet periods, excluding updates related to the press conferences. Here, the prior coefficient is expected to be smaller than during the quiet period, as the ECB communicates with the public during this time. The third specification restricts the sample to only updates to press conferences by enforcing two requirements: A posterior must be posted within 24 hours of the monetary policy announcement on a press conference day, and a prior must be posted in the most recent quiet period. The former ensures that the response is indeed due to information from the press conference. The latter ensures that the update does not include information from other ECB communication. Whether updating around press conferences it is larger or smaller than outside quiet periods depends on whether the biggest shift in beliefs come from predominantly monetary policy related news, or other types of news that relate to the ECB (e.g., announcements of personnel changes).

The main results only focus on updates to press conferences. Instead of press conference fixed effects, I now include actual values for each press conference's signal $j \in [1, \dots, J = 48]$ that allow to study how much individuals use novel information from the signal compared to their prior beliefs when updating their beliefs. To disentangle ingroup-driven effects on the signal, I also interact the signal with the ingroup dummy:

$$Posterior_{i,t} = \beta_1 Prior_{i,t} + \beta_2 Prior_{i,t} * Ingroup_{i,t} + \beta_3 Ingroup_{i,t} + \beta_4 Signal_t + \beta_5 Signal_t * Ingroup_{i,t} + \epsilon_{i,t} \quad (A.3)$$

This allows to identify how much ingroup and outgroup individuals use the signal.

Results I run pooled regressions with all individuals by estimating Equation (A.2) using OLS to show how beliefs are updated at different points in time: during quiet periods, outside quiet periods, and around press conferences.² Beliefs are updated the least during quiet periods (see column (1) in Table A.6), more around press conferences (see column (3)), and most during the 5 weeks outside quiet periods (see column (2)). This makes sense intuitively, as the news about central banking are by definition limited during the quiet period, thus I expect posteriors to resemble their priors strongly. By contrast, press conferences communicate the most important monetary policy decisions, which is why seeing a larger shift in beliefs between the preceding quiet period and right after the press conference speaks for the credibility of any insights from analyses of these observational data. The biggest shift in beliefs occurs during the 5 weeks after a press conference and before the quiet period starts. During these weeks, many different kinds of news are released, which might explain the larger change in sentiment than when news are mostly about monetary policy.³

Across all three time periods, there exist ingroup effects, whereby ingroup individuals rely less on their priors (by around a third) than outgroup individuals (see Table A.6). Lower prior reliance suggests higher responsiveness to new information; therefore, signals from the central bank are more effective in shaping beliefs.

Table A.7 shows belief updating around the press conference to better understand how individuals use their priors compared to the signal they receive. Column (1) includes fixed effects of individual press conferences (see Equation (A.2)). While I do not obtain a coefficient for the signal this way, this specification requires fewer assumptions about the exact signals that individuals receive. The prior coefficient for outgroup individuals is at 0.49***, compared to ingroup individuals for whom the prior coefficient is at 0.29***. This may be interpreted as ingroup individuals relying almost a third less on their prior beliefs—thus being a third more open to novel information—when forming their beliefs following a press conference. This is even more pronounced when including values for the signals: Outgroup individuals' reliance on their priors is at 0.84***, while this value is only 0.47*** for ingroup individuals (more than 40% less). Regressions shown in columns (2)-(5) are again estimated by OLS but now do not include time fixed effects (see Equation (A.3)), but instead the actual signal of each press conference. Columns (4) and (5) are the same regression as in column (3), but with the sample limited to non-experts and experts, respectively. Even when accounting for the actual signal that agents receive, ingroup effects are extremely persistent across specifications and subsamples. Ingroup agents hold on to their priors about a third less (even though both in- and outgroup agents get a larger coefficient for their posteriors compared to the regression specification without signal values). And similarly, ingroup agents use the signal more than twice as much as outgroup agents. In fact, their signal use becomes similar in size to their prior reliance. Figure A.5 visualizes these results for the entire sample, and the non-(expert) subsamples.

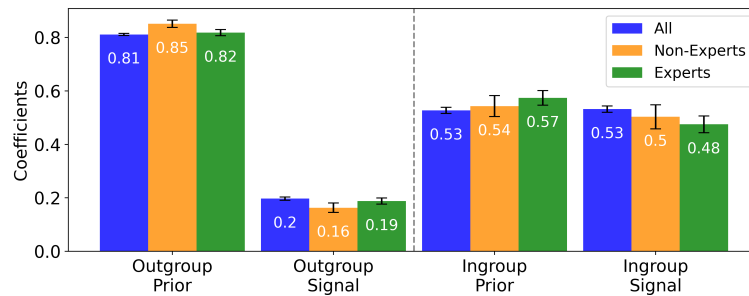
²Regression results estimating individual coefficients from separate regressions for each agent confirm pooled regression findings and are available upon request.

³This might be even less surprising considering that news are expected at press conferences, and the timing of these press conferences is not a surprise.

Table A.6. Belief Updates at Different Points in Time

	<i>Quiet Period (QP)</i>	<i>Outside QP (NQP)</i>	<i>Press Conferences (PCs)</i>
<i>Dep. Var.: Posterior</i>	(1)	(2)	(3)
<i>Prior</i>	0.511*** (0.002)	0.339*** (0.001)	0.492*** (0.004)
<i>Ingroup*Prior</i>	-0.403*** (0.005)	-0.285*** (0.002)	-0.455*** (0.010)
<i>Ingroup</i>	0.224*** (0.003)	0.154*** (0.001)	0.255*** (0.005)
<i>Outgroup Prior</i>	0.511*** (0.002)	0.339*** (0.001)	0.492*** (0.004)
<i>Ingroup Prior</i>	0.331*** (0.002)	0.208*** (0.001)	0.292*** (0.005)
<i>Fixed Effects</i>	QP	NQP	PC
<i>N</i>	228,733	2,009,176	63,757
<i>R-squared</i>	0.958	0.958	0.954
<i>Posterior Variable</i>	most recent in same QP	most recent in same NQP	most recent from previous QP
<i>Prior Variable</i>	any in QP	any outside QP	within 24hrs after PC (closest to PC)

Notes: Regressions are estimated using OLS. Standard errors are shown in parentheses. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level. N refers to the number of tweets. All regressions use individual tweets. The belief variables (prior and posterior) are determined by dictionary-based tweet sentiment. QP refers to “Quiet Period” and NQP to “Not Quiet Period” i.e., outside quiet periods. The quiet period refers to the 7 days leading up to a press conference, where there should not be any central bank signals (European Central Bank, 2016). To be included in the (N)QP regressions of columns (1) and (2) a user must tweet at least twice within the same (N)QP i.e., uninterrupted by a press conference (PC). To be included in the regression of column (3) a user must tweet in a quiet period (the prior) and within 24 hours of the following press conference, where the first tweet within this time window is selected as the posterior. Values for the “Ingroup Prior” are computed by taking the linear combination of relevant regression coefficients.

Figure A.5. Belief Updating after Press Conferences By Expertise

Notes: Bars show summed up OLS regression coefficients (base + interaction) with corresponding 95% confidence intervals of the priors (signals) in the left (right) panel for either individuals of the out- or the ingroup. The blue bars shows results for the entire sample, and the orange (green) bars for the subsample of non-experts (experts). The (non-)expert classification follows Ehrmann and Wabitsch (2022).

Table A.7. Belief Updating after Press Conferences

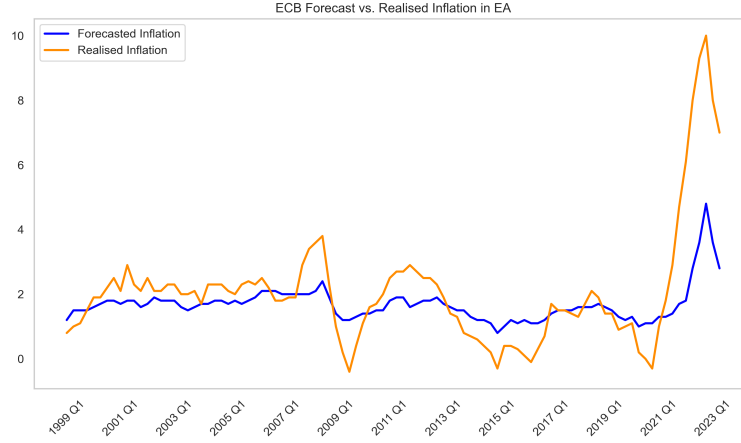
<i>Dep. Var.: Posterior</i>	(1)	(2)	(3)	(4)	(5)
	<i>All</i>		<i>Non-Experts</i>		<i>Experts</i>
<i>Prior</i>	0.492*** (0.004)	0.835*** (0.002)	0.811*** (0.002)	0.851*** (0.007)	0.818*** (0.006)
<i>Ingroup*Prior</i>	-0.455*** (0.010)	-0.806*** (0.010)	-0.782*** (0.010)	-0.795*** (0.037)	-0.701*** (0.026)
<i>Signal</i>		0.166*** (0.003)	0.197*** (0.003)	0.163*** (0.009)	0.188*** (0.006)
<i>Ingroup*Signal</i>			-0.164*** (0.003)	-0.147*** (0.022)	-0.169*** (0.016)
<i>Ingroup</i>	0.255*** (0.005)	0.440*** (0.006)	0.498*** (0.006)	0.487*** (0.021)	0.457*** (0.015)
<i>Outgroup Prior</i>	0.492*** (0.004)	0.835*** (0.002)	0.811*** (0.002)	0.851*** (0.007)	0.818*** (0.006)
<i>Ingroup Prior</i>	0.292*** (0.005)	0.469*** (0.005)	0.527*** (0.006)	0.543*** (0.020)	0.574*** (0.014)
<i>Outgroup Signal</i>			0.197*** (0.003)	0.163*** (0.009)	0.188*** (0.006)
<i>Ingroup Signal</i>			0.532*** (0.006)	0.503*** (0.023)	0.475*** (0.016)
<i>Fixed Effects</i>	PC	-	-	-	-
<i>N</i>	63,757	63,757	63,757	5,509	11,478
<i>R-squared</i>	0.954	0.946	0.946	0.944	0.957

Notes: Regressions are estimated using OLS. Regressions in columns (2)–(5) are estimated with no constant. Standard errors are shown in parentheses. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level. N refers to the number of tweets. All regressions use individual tweets. The belief variables (prior and posterior) are determined by dictionary-based tweet sentiment. The signal is based on the OIS-2Y volatility indicator. Press Conference (PC) fixed effects refer the inclusion of dummies for each 6-week period starting at the moment of the press conference’s monetary policy announcement. To be included in the regressions of columns (1)–(5) a user must tweet in a quiet period (the prior) and within 24 hours of the following press conference, where the first tweet within this time window is selected as the posterior. Columns (4) and (5) show the regression as shown in column (3), but restrict the sample to either non-experts or experts only. The (non-)expert classification follows the benchmark in Ehrmann and Wabitsch (2022). Values for the “Ingroup Prior” and “Ingroup Signal” are computed by taking the linear combination of relevant regression coefficients.

A.2 Additional Information on the Experiment

Data in Forecasting Tasks The figure below shows the data used in the experiment for inflation and its forecasts.

Figure A.6. Quarterly EA Inflation, SPF Forecasts, and SPF Forecast Errors



Notes: The figure shows realized quarterly EA HICP and forecasts of quarterly EA inflation (1-year-ahead). Data for inflation are taken from the ECB's Statistical Data Warehouse. Data for the inflation forecasts are based on the HICP forecast in the ECB Survey of Professional Forecasters (SPF), which can be found on the ECB website.

Table A.8. Overview of Randomly Selected Sequences

10-Period Starting Date	10-Period Average of Forecasting Precision	10-Period Average of Inflation	10-Period STD of Inflation	Next-Period Signal	Next-Period Realisation
2006Q3	1.6	2.6	0.8	1.4	1
2008Q1	1.1	1.7	1.5	1.5	1.7
2004Q1	2.4	2.2	0.2	2.1	2.2
2011Q4	1.7	1.9	0.8	1.2	0.6
2001Q2	1.9	2.3	0.3	1.6	2.1
2017Q3	4.8	1.5	0.4	1.3	1.1

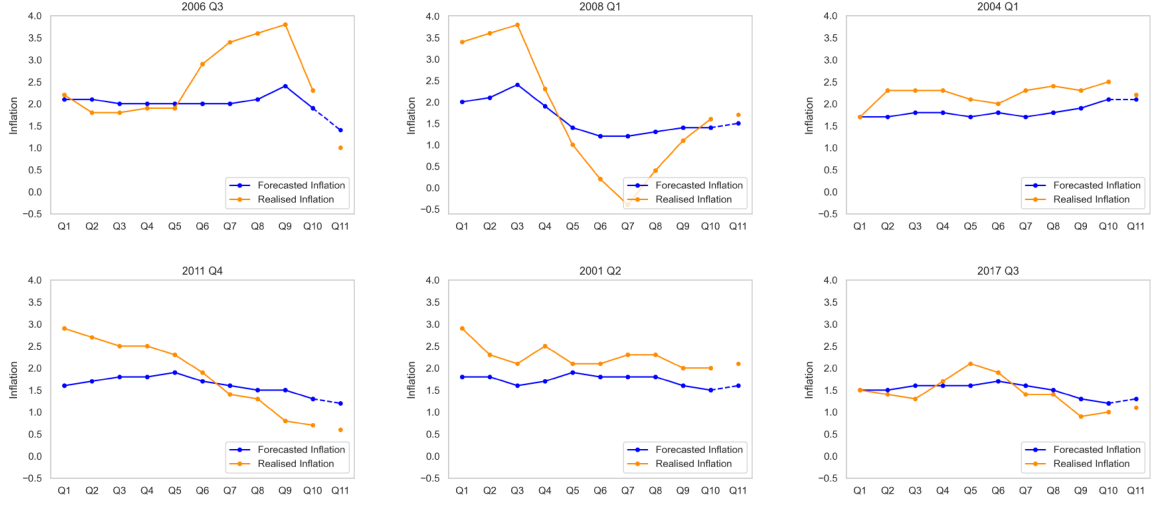
Notes: The table shows 10-period forecasting precision averages, realized inflation averages, and standard deviations of realized inflation for 6 randomly selected sequences. The indicated date marks the beginning of a 10-quarter sequence. Next-Period Signal and Realization refer to the forecast and realization of inflation, respectively, in the 11th quarter of a given sequence. Data for realized EA inflation and for SPF's inflation forecasts are taken from the ECB's Statistical Data Warehouse.

Details on Forecast Incentives Point forecasts for the prior and the posterior are incentivized with a bonus payment that is based on a participant's forecasting score $F_{i,t}$:

$$F_{i,t} = 6 * 3^{-|\mathbb{E}_{i,t-1}\{\pi_t\} - \pi_t|}, \quad (\text{A.4})$$

where π_t is inflation at period t and $\mathbb{E}_{i,t-1}\{\pi_t\}$ is its forecast. A perfect forecast yields $F_{i,t} = 6$. This forecasting score is reduced by 2/3 of its value for each percentage point increase in the forecast error.

Figure A.7. Inflation Sequences



Notes: Sequences are randomly selected and based on real EA HICP and SPF 1-year-ahead forecasting data. The solid line represents the inflation (forecast) history that participants are shown initially. The dashed “predicted” line leads to the signal participants receive, and the unconnected final dot of the “realized” series is the value participants try to forecast. The first quarter of each sequence can be found at the top of each chart, however, note that date information will never be given to participants. Sequences are randomized across all messenger treatments.

Range forecasts are incentivized using the scoring rule below, which follows the logic of e.g., Pfajfar and Žakelj (2016), Rholes and Petersen (2021) and Petersen and Rholes (2022).

$$U_{i,t}(r_{i,t}) = \begin{cases} 0 & \pi_{i,t} \notin \left[\frac{u_{i,t}}{1+r_{i,t}}, \overline{u_{i,t}} \right] \\ \phi \left(\frac{1}{1+r_{i,t}} \right) & \pi_{i,t} \in \left[\frac{u_{i,t}}{1+r_{i,t}}, \overline{u_{i,t}} \right] \end{cases}, \quad (\text{A.5})$$

where ϕ is a scalar that can be adjusted to scale average earnings, which are strictly increasing in ϕ . For this experiment, I set ϕ to 6. $u_{i,t}$ represents the lower bound of a participant’s forecast uncertainty, $\overline{u_{i,t}}$ the upper bound, and $r_{i,t} = \frac{|\overline{u_{i,t}} - u_{i,t}|}{u_{i,t}}$ represents a participant’s forecast uncertainty. The intuition of this scoring rule is that if realized inflation falls out of a participant’s forecast range, a participant earns nothing. But if instead realized inflation does fall within a participant’s uncertainty bounds, the participant earns a payoff that is decreasing in the magnitude of her uncertainty. Thus, participants are incentivized to keep this range as narrow as possible, but also to ensure that it covers any inflation value they view as probable.

After submitting their forecasts, participants are not immediately shown the realized inflation values. This approach avoids learning effects between forecasting tasks and ensures that their performance does not influence responses to subsequent questions about trust in institutions or policymakers similar to treatments. Instead, participants are informed of their point and range forecasting performance across all 12 forecasts (6 prior and 6 posterior forecasts) at the end of the survey. This is also when they learn which of the 24 decisions⁴

⁴12 point and 12 range forecasts across 6 initial and 6 updated forecasts

has been selected randomly for their bonus payment. Randomly selecting one decision ensures that participants are incentivized to put high effort into each decision, as they cannot tell which of their forecasts will matter for their final payment.

Details on Treatments and Randomization Procedures In the *Generic Expert Treatments* (related to H1), participants encounter generic experts of four possible nationalities (Germany, France, Italy, and Spain), one of whom matches the participant’s nationality (ingroup) and another randomly selected from the other three nationalities (outgroup).

The *ECB Expert Treatments* (related to H2 and H3) introduce ECB-affiliated experts with specified nationalities. For consistency, the outgroup expert in this treatment matches the outgroup nationality used in the Generic Expert Treatments, allowing a clearer comparison between ingroup and outgroup messengers across institutional and non-institutional contexts.

The *Institutional Expert Treatments* (related to H3 and H4) feature messengers from either the ECB or the participant’s national central bank (NCB), such as the Deutsche Bundesbank or Banque de France, without specified nationalities. This treatment aims to test the effectiveness of delegating communication to NCBs, and whether homophily or heterophobia dominates ingroup effects with institutional context.

To minimize fatigue, each participant encounters only six out of the ten treatments—two from each group—with treatment order and inflation scenarios randomized to prevent learning effects and ensure valid comparisons.

Details on the Post-Experimental Survey To be able to account for any pre-existing perceptions of institutions and policymakers that could influence decision-making, participants are surveyed on monetary institutions (the ECB and the respective NCB) and policymakers that are representative of the messenger treatments 5–10. Participants are asked to report their *trust*⁵ in these institutions and policymakers, and their respective *exposure* (i.e., the extent to which they know of and follow news about these institutions and policymakers). As representative ECB experts of the in- and outgroup, I take advantage of the fact that all four nationalities are represented among the six ECB board members at the time of the experiment.⁶ As further in- and outgroup policymakers, NCB governors are included in the survey. The NCB governors at the time of the experiment are: François Villeroy de Galhau for France, Joachim Nagel for Germany, Ignazio Visco for Italy, and Pablo Hernández de Cos for Spain. Table A.10 provides an overview of chosen representative policymakers.

⁵Trust is only elicited for policymakers and institutions that the participant indicates to know of, and answers are given on a 7-point Likert scale (see for details Figure A.24).

⁶“Christine Lagarde” is used to proxy for the French ingroup ECB expert, and for all others to proxy an outgroup ECB expert. Similarly, “Luis de Guindos” proxies the Spanish, “Isabel Schnabel” the German, and “Fabio Panetta” the Italian ECB ingroup expert. And respectively, these policymakers proxy an outgroup ECB expert, whenever the participant nationality does not match the policymaker’s nationality.

Information Table A.9 shows all information pieces revealed by clicking “Read more”.

Table A.9. Additional Information Pieces and Corresponding Inflation Scenarios

Inflation Scenario	Information Pieces
2006 Q3	1) Overall, the evidence that inflationary pressures are diminishing further has increased and, looking forward, inflation rates are expected to be in line with price stability, supporting the purchasing power of incomes and savings. 2) The recent decline in headline inflation mainly reflects the considerable easing in global commodity prices over the past few months, which more than offsets the impact of the sharp rise in unit labour costs in the first half of this year. Lower commodity prices and weakening demand should lead to inflationary pressures diminishing further. 3) The large declines in commodity prices and the impact of weakening demand on price developments strongly impacted this forecasts.
2008 Q1	1) Price developments are expected to remain moderate over the medium-term horizon. Global inflationary pressures may persist, while domestic price pressures are expected to remain low. 2) The still weak annual growth rate of bank loans to the private sector conceals the fact that monthly flows have now been positive for a number of months. At the same time, these aggregate developments continue to reflect mainly an ongoing strengthening in the annual growth of loans to households, while the annual growth of loans to non-financial corporations has remained negative. 3) Overall, price stability is expected to be maintained over the medium term, thereby supporting the purchasing power of Euro area households. Inflation expectations remain firmly anchored.
2004 Q1	1) The inflation rates are forecasted to remain elevated. Factors contributing to this include higher energy prices, indirect effects of past oil price increases, and potential stronger-than-expected wage developments. 2) The strength of monetary and credit growth, coupled with ample liquidity, pose inflationary risks over the medium to longer term. 3) The rapid rate of monetary growth is primarily driven by the stimulative impact of low-interest rates.
2011 Q4	1) The risks to the economic outlook for the Euro area are considered to be broadly balanced over the medium term, both for upside and downside risks to price developments. 2) The monetary analysis confirms subdued underlying growth in broad money and credit, and the annual growth rate of loans to the private sector continues to contract. 3) Progress has been made in improving the funding situation of banks, but further steps are needed to strengthen the resilience of banks and reduce the fragmentation of Euro area credit markets.
2001 Q2	1) Real economic activity in the Euro area was weak recently, but survey data and recent financial market developments indicate a gradual upturn in economic activity. 2) The expected pick-up in activity is supported by both external factors, such as global recovery, and domestic factors, such as ongoing adjustment efforts by companies to enhance competitiveness and profitability. 3) Downside risks to the main scenario for economic growth have declined, but they have not disappeared. Macroeconomic imbalances in some regions and high oil prices are mentioned as potential risks to economic activity in the Euro area.
2017 Q3	1) While there are some initial signs of stabilization in the growth slowdown and a mild increase in underlying inflation, overall inflation remains low. 2) Ongoing employment growth and increasing wages are underpinning the resilience of the Euro area economy. 3) Inflation expectations are at low levels, and measures of underlying inflation have remained generally muted. Inflation is expected to increase over the medium term, supported by monetary policy measures, economic expansion, and solid wage growth.

Notes: Information pieces are linked to inflation scenarios (each scenario’s first quarter shown in the first column). Each inflation scenario includes three buttons, each displaying one piece of information. Information pieces are summarized statements based on ECB press conferences, which can be found in full [here](#).

Table A.10. Representative Policymakers by Nationality

	ECB Board Member	NCB Governor
France	Christine Lagarde	François Villeroy de Galhau
Italy	Fabio Panetta	Ignazio Visco
Germany	Isabel Schnabel	Joachim Nagel
Spain	Luis de Guindos	Pablo Hernández de Cos

Notes: Overview of all representative in- and outgroup policymakers by nationality used in post-experimental survey. ECB Board Members represent in- and outgroup ECB experts: “Christine Lagarde” represents the French ingroup ECB expert, “Luis de Guindos” proxies the Spanish, “Isabel Schnabel” the German, and “Fabio Panetta” the Italian ECB ingroup expert. Similarly, the right column presents all four NCB governors at the time of the experiment. All participants face all policymakers, which are counted as ingroup (outgroup) policymakers whenever nationalities of participants and policymakers match (mismatch).

A.2.1 Power Analysis

Power analysis is commonly used to determine the required number of observations to preclude insignificant results that are due to low statistical power (see e.g., Cohen (2013)). My choice of sample size for the experiment follows from such power analysis for comparisons of dependent means (since I use a within-subjects design). Based on the observational results from my Twitter analysis, pilot data, and findings from a comparable study (i.e., D’Acunto et al. (2022)), the expected effect size is approximately 0.1, ranging between 0.1 and 0.3. Figure A.8 visualizes the required sample size for each of these expected effect sizes at various levels of required power and significance. The most commonly used power (alpha) specification is 0.8 (0.05), which is represented by the dark blue line. Considering this specification, my ideal sample size is just below 800, as this would allow me to statistically identify effect sizes of 0.1. However, due to budget constraints, the sample size of the experiment is limited to 400 participants (100 per nationality), which allows me to safely identify effect sizes of approximately 0.14 and above. This means that if the observed effects are smaller and statistically insignificant, their lack of significance might be due to insufficient sample size.

A.2.2 Bayesian Belief Updating Framework

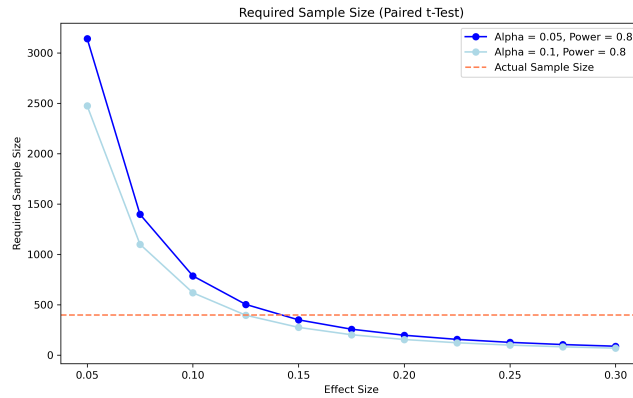
To assess how agents update their inflation expectations following central bank communication, I use a standard Bayesian belief updating framework. This framework estimates how agents balance novel information (central bank communication) against their existing beliefs (priors). In the Bayesian case, agents weigh information and priors equally when both have the same precision (e.g., Veldkamp (2011)).

Assume macroeconomic expectations are normally distributed, where agent i believes inflation x has a mean of A_i and precision α_i . The prior belief is $x \sim \mathcal{N}(A_i, \alpha_i^{-1})$, and agents receive a signal B about inflation x from the central bank: $B = x + e$, $e \sim \mathcal{N}(0, \beta^{-1})$, where the signal has precision β . Applying Bayes’ law, agents update their beliefs as:

$$E_i[x|B] = \frac{\alpha_i A_i + \beta B}{\alpha_i + \beta}, \quad (\text{A.6})$$

which shows that the posterior belief is a weighted average of the prior and the signal,

Figure A.8. Power Analysis



Notes: The figure shows optimal sample size to ensure significance of effects for various power and alpha levels for expected effect sizes (paired t-test). The red horizontal dashed line indicates the actual sample size that was achievable with the available funding.

based on their relative precision.

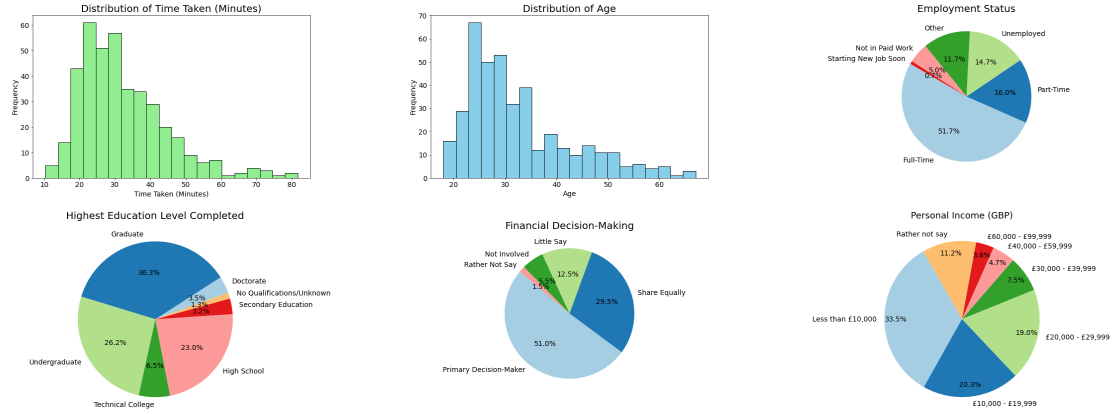
A.2.3 Additional Experimental Results, Tables and Figures

Participants in the Experiment The following background data of participants is provided by Prolific (as defined before running the study): age, sex, income bracket, marital status, employment status, student status, ethnicity, country of birth/residence, years lived in current country of residence, country spent most time in before turning 18, extent of financial decision-making, primary and fluent languages, highest obtained education, migration history, investments, property ownership, information on whether participant has lived abroad, is mono/multicultural, or has been raised monolingual. Figure A.9 shows some of the participants' characteristics descriptively in more detail.

While the sample does not attempt to be representative of the populations in Germany, Spain, Italy or France, the participants in the collected sample are much more diverse compared to common samples in lab experiments, as these commonly include students only. Participants' age ranges from 18 to 67 and averages at 33. More than half of the participants are employed full-time, only around 35% are students, more than half of participants earn more than 10,000 GBP per year, more than half of the participants are married or in a relationship, around a third of the sample is female (compared to two thirds being male), and the highest level of education obtained is distributed fairly evenly among high school, undergraduate, and graduate levels. Most participants complete the experiment in half an hour. A few outliers take around an hour. Prolific automatically disqualifies participants from submitting the experiment if they take longer than 90 minutes.

Average Signal Use The table below shows that participants under-use the signal compared to a Bayesian benchmark across treatments.

Figure A.9. Descriptive Evidence of Experiment Participants



Notes: The figure shows descriptively various characteristics of experiment participants, where the characteristic of interest is shown on top of each panel. Only the histogram on the top left is not a participant characteristic but shows the amount of time (in minutes) participants needed to complete the experiment.

Table A.11. Regression Results By Nationality

	(1) Overall	(2) Germany	(3) Spain	(4) France	(5) Italy
Signal	0.899*** (0.012)	0.881*** (0.025)	0.889*** (0.025)	0.877*** (0.024)	0.941*** (0.026)
Prior	1.206*** (0.012)	1.234*** (0.025)	1.198*** (0.022)	1.262*** (0.025)	1.139*** (0.025)
R-squared	0.961	0.961	0.963	0.964	0.959
N	2,385	600	594	593	598

Notes: The table shows the regression results of estimating Equation 1.1 by OLS without a constant. Standard errors are shown in brackets. Results for the entire sample are shown in column (1), while the other columns show results for samples split by nationality. Stars correspond to the following p-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Mechanism The table below shows the role of perceived signal quality and trust.

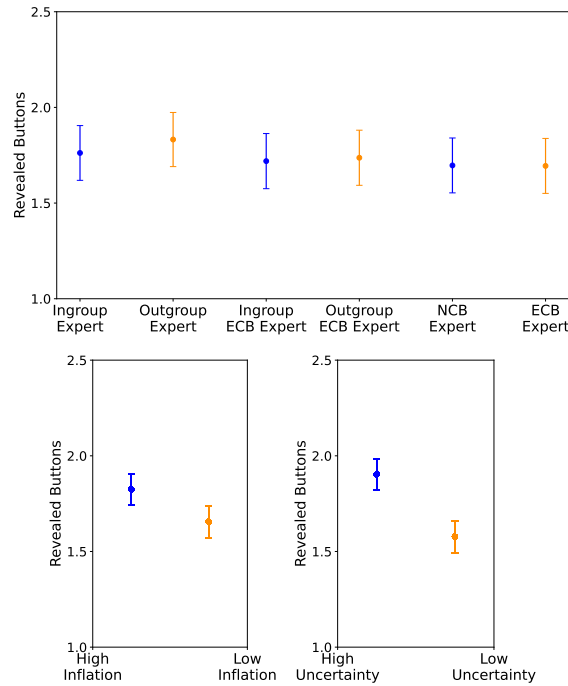
Table A.12. Mechanism: Perceived Quality and Trust

	(1)	(2)	(3)	(4)
<i>Pure Ingroup Effect (H1):</i>	0.052*** (0.017)	0.010** (0.005)		
R-squared	0.994	0.992		
N	795	795		
<i>Ingroup Effect for ECB Experts (H2):</i>	0.028* (0.017)	0.005 (0.004)	0.053 (0.044)	0.006 (0.015)
R-squared	0.994	0.993	0.998	0.997
N	795	795	315	315
<i>Homophily - Ingroup ECB Expert vs. Generic ECB Expert (H3):</i>	0.035** (0.018)	0.009** (0.004)	0.051** (0.022)	0.021*** (0.007)
R-squared	0.993	0.992	0.996	0.995
N	794	794	612	612
<i>Heterophobia - Outgroup ECB Expert vs. Generic ECB Expert (H3):</i>	0.013 (0.017)	0.008** (0.004)	-0.004 (0.033)	0.012 (0.011)
R-squared	0.994	0.993	0.998	0.998
N	795	795	477	477
<i>NCB vs ECB: Institutions Effect (H4):</i>	0.034** (0.017)	0.008** (0.004)	0.033* (0.017)	0.005 (0.004)
R-squared	0.994	0.994	0.994	0.993
N	795	795	779	779
Inflation Scenario	✓	✓	✓	✓
Individual-FE	✓	✓	✓	✓
Perceived Messenger Ability		✓		
Trust				✓
Knowing PMs/Inst			Yes	Yes

Notes: Effects result from comparing coefficients of interest of the same regression. *Inflation Scenario* refers to controlling for the underlying data of forecasting tasks and the order in which they appeared. *Perceived Messenger Ability* refers to the self-reported 7-point Likert scale indication of how able each messenger is thought to be at forecasting inflation and providing economic analyses. *Knowing PMs/Inst* refers to whether participants indicated to at least have heard of the institutions or real policymakers (see Section 1.3.1 for details on this) matching the messenger treatments. Conditional on knowing representative PMs and/or institutions, self-reported *Trust* is available (based on 7-point Likert scale). *N* refers to the number of observations (forecasting tasks) in the regression. *N* may slightly vary across treatments due to instances of infinite prior precision. Column (1) shows the main results again. Column (2) interacts signal use with perceived ability. Column (4) interacts this effect with trust in a representative policymaker or in the institution. Column (3) mirrors column (1) but on the restricted sample where policymakers and institutions are known in each treatment, which allows for better comparison with results in column (4). Stars correspond to the following p-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Attention to Information The top panel in Figure A.10 shows that while attention to information across messenger treatments is fairly constant (around 1.8 revealed information buttons), the bottom panel of the figure shows that attention indeed responds to both the level of inflation and its uncertainty.

Figure A.10. Requesting Additional Information Pieces (With 95%-CI)



Notes: Top row shows the additional information pieces requested via the average number of buttons clicked (with 95%-CI) by messenger treatments. Bottom row shows the same by inflation level (left panel) or inflation uncertainty (right panel). Out of the six inflation scenarios, the three scenarios with the highest (lowest) inflation shown as “High (Low) Inflation”. Similarly, the top (bottom) three scenarios in terms of the inflation history’s standard deviation are shown as “High (Low) Inflation Uncertainty”.

Attention and Ingroup Effects When participants request additional information, they show stronger ingroup effects across hypotheses. I attribute this to these participants being attentive. In fact, ingroup effects disappear when limiting the sample to forecasting tasks where no buttons are revealed (i.e., “no attention”), highlighting that treatments might not have been salient to these inattentive participants. In turn, effects become stronger the more attention individuals pay (i.e., the more buttons they reveal). Note that since attention is not exogenous in my experiment, these effects are not strictly causal. For instance, a lack of attention could also represent low effort exertion, resulting in missing the treatment variation.

Two robustness exercises show that it is indeed attention—not the additional information—that drives ingroup effects on signal use. Column (2) in Table A.13 compares only forecasting tasks where all information is retrieved and confirms effects (except that the ingroup effect within ECB context loses significance). Similarly, column (3) confirms robustness of all effects with the full sample, controlling for revealed buttons (for given inflation scenarios).

Table A.13. The Role of Attention and Being Informed

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Pure Ingroup Effect (H1):</i>	0.052*** (0.017)	0.066*** (0.025)	0.056*** (0.018)	0.057** (0.024)	0.039 (0.036)			
R-squared	0.994	0.995	0.994	0.995	0.995			
N	795	463	795	488	307			
<i>Ingroup Effect for ECB Experts (H2):</i>	0.028* (0.017)	0.029 (0.025)	0.029* (0.017)	0.018 (0.027)	0.033 (0.035)	0.053 (0.044)	-0.006 (0.064)	-0.037 (0.107)
R-squared	0.994	0.996	0.994	0.995	0.995	0.998	0.999	0.999
N	795	446	795	465	330	315	193	122
<i>Homophily - Ingroup ECB Expert vs. Generic ECB Expert (H3):</i>	0.035** (0.018)	0.083*** (0.026)	0.040** (0.018)	0.081*** (0.025)	-0.007 (0.032)	0.051** (0.022)	0.110*** (0.031)	-0.030 (0.042)
R-squared	0.993	0.995	0.994	0.995	0.995	0.996	0.997	0.998
N	794	437	794	462	332	612	361	251
<i>Heterophobia - Outgroup ECB Expert vs. Generic ECB Expert (H3):</i>	0.013 (0.017)	0.059** (0.023)	0.007 (0.017)	0.057** (0.022)	-0.032 (0.036)	-0.004 (0.033)	0.047 (0.040)	-0.251** (0.097)
R-squared	0.994	0.996	0.994	0.996	0.995	0.998	0.999	1.000
N	795	441	795	465	330	477	284	193
<i>NCB vs ECB: Institutions Effect (H4):</i>	0.034** (0.017)	0.058** (0.025)	0.038** (0.018)	0.057** (0.024)	0.005 (0.034)	0.033* (0.017)	0.052** (0.024)	0.006 (0.035)
R-squared	0.994	0.995	0.994	0.995	0.995	0.994	0.996	0.995
N	795	433	795	464	331	779	455	324
Inflation Scenario	✓	✓	✓	✓	✓	✓	✓	✓
Individual-FE	✓	✓	✓	✓	✓	✓	✓	✓
Attention		Full	Controlled	Yes	No		Yes	No
Knowing PMs/Inst						Yes	Yes	Yes

Notes: Effects result from comparing coefficients of interest of the same regression. *Inflation Scenario* refers to controlling for the underlying data of forecasting tasks and the order in which they appear. *Attention - Full* refers to the subsample of inflation tasks, in which participants click on all information buttons to reveal additional information. *Attention - Controlled* refers to the full sample of inflation tasks, but controls for which buttons participants click on (for a given inflation scenario) to reveal additional information. *Attention - Yes* refers to the subsample of inflation tasks, in which participants click on at least one button to reveal additional information. Conversely, *Attention - No* reflects the subsample of inflation tasks, for which not a single button to reveal more information is clicked on. *Knowing PMs/Inst* refers to whether participants indicate to at least have heard of the institutions or real policymakers matching the messenger treatments. *N* refers to the number of observations (forecasting tasks) in the regression. *N* may slightly vary across treatments due to instances of infinite prior precision. Stars correspond to the following p-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Interestingly, attention further highlights another potential explanation of ingroup effects within the institutional context. When representative ECB policymakers are known but attention is low, there is substantial heterophobia (-0.251^{**}) This suggests that nationality may be used as a shortcut. Conversely, homophily is very large (0.110^{***}) when ECB policymakers are known and participants are attentive (see columns (6)-(8) in Table A.13).

A.2.4 Information Availability: Exposure to News in the Media

The survey accompanying the experiment highlights that participants receive more news about in- than outgroup policymakers. The probabilities of having heard of, knowing, and following news about policymakers⁷ are all larger for ingroup policymakers (see Table A.14).

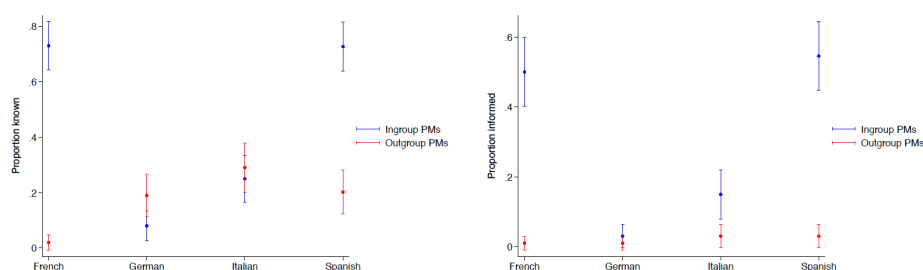
Table A.14. Information Availability and Reach: Survey Evidence

	(1) Heard Of	(2) Know	(3) Informed
Ingroup=1	0.080** (0.035)	0.271*** (0.031)	0.286*** (0.024)
Constant	0.489*** (0.025)	0.175*** (0.022)	0.020 (0.017)
R-squared	0.01	0.09	0.15
N	798	798	798

Notes: The representative in- and outgroup policymakers are all ECB board members (Lagarde, de Guindos, Schnabel, Panetta) but regressions for NCB governors (Villeroy de Galhau, de Cos, Nagel, Visco) are similar. Regressions are estimated by OLS. The Ingroup dummy indicates the respective probabilities for ingroup compared to outgroup policymakers. Data come from the survey question shown in Figure A.23.

The likelihood of knowing policymakers increases by roughly a third if they are in the ingroup rather than outgroup. Participants are 29% more likely to follow news of a given policymaker if they are in the ingroup. Figure A.11 plots the shares of respondents who indicate knowing or following news about in- and outgroup policymakers by participant nationality. Effects for ECB ingroup policymakers seem to be driven by the French and Spanish—coinciding with the nationalities of the ECB president and vice president. This suggests that nationality strongly matters, especially for these two ECB Board positions.

Figure A.11. Knowing Representative Policymakers and Receiving News



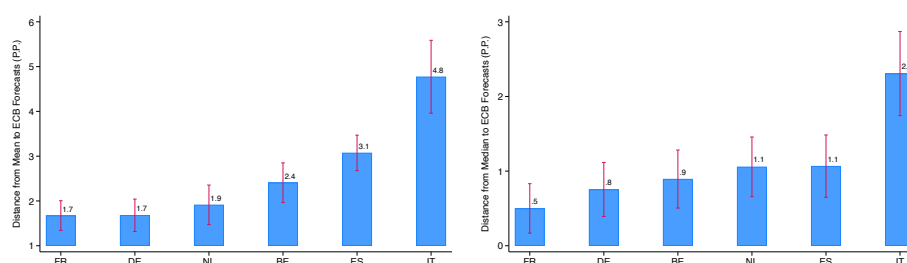
Notes: The left (right) panel shows the share of participants in the experiment who report knowing (receiving news about) representative in- and outgroup policymakers (PMs). The average response across all three outgroup PMs is used. These plots show results only for ECB board members (Lagarde, de Guindos, Schnabel, Panetta) but plots for NCB governors (Villeroy de Galhau, de Cos, Nagel, Visco) are very similar.

⁷As described earlier in the paper, the policymakers used to proxy for messengers are NCB governors (Villeroy de Galhau, de Cos, Nagel, Visco) and ECB board members (Lagarde, de Guindos, Schnabel, Panetta)

A.2.5 Real-World Forecasting

The experimental findings suggest that ingroup agents update more strongly to a signal; thus their inflation forecasts should—*ceteris paribus*—more closely resemble the forecasts by the ECB. To check whether this is indeed the case, I compare the distance between the ECB’s 12 months-ahead inflation forecasts to forecasts of households in the EA. Values for the former are taken from the ECB’s Macroeconomic Projection Database (MPD). The MPD is published four times a year (in March, June, September and December).⁸ The latter come from the ECB’s Consumer Expectations Survey (CES), which is a representative survey of consumers in selected EA countries started in 2020. I use their aggregated monthly mean and median indicators of quantitative household inflation expectations over the next 12 months.⁹

Figure A.12. Comparing Household Inflation Forecasts to ECB Projections



Notes: The left (right) panel shows the difference between mean (median) household twelve months-ahead inflation forecasts and ECB projection (lagged by 1 month) with 95% confidence intervals. I consider fifteen survey waves for six countries, whereby data for Belgium are missing for the first wave (April 2020).

Figure A.12 shows that forecasts of 1 year-ahead inflation by households in France indeed most closely resembled ECB’s forecasts (lagged by 1 month).¹⁰ Available data spans from April 2020 until December 2023, when the ECB’s president was French (Christine Lagarde). During this time, the mean forecasts for French households are on average 1.7 percentage points above the ECB’s projections, the lowest distance across countries (together with German households). Across countries, mean household forecasts are on average 2.6 percentage points above ECB forecasts. Similarly, French median household forecasts are most similar to ECB forecasts (0.5 percentage point distance), with a cross-country average distance of 1.1 percentage points. While these findings are correlations rather than causal effects, the trend is in line with ingroup agents updating more closely to the signal.

⁸The MPD contains information on the outlook for the EA and contributes to the ECB Governing Council’s assessment of economic developments and risks to price stability. The June and December projections are conducted by Eurosystem staff, while the March and September projections are conducted by ECB staff.

⁹The survey provides mean and median values that are population-weighted and winsorised at the 2nd and 98th percentiles of the weighted distribution of responses for each survey wave and country.

¹⁰Not lagging ECB’s forecast results in very similar findings.

A.3 Interface of the Experiment

Figure A.13. Welcome Page

Welcome!

Name of Study: Forecasting Inflation

Description: The aim of this experiment is to better understand macroeconomic expectations. No prior knowledge is necessary to participate in the experiment. Your task will be to forecast inflation in different economic scenarios. After these forecasting tasks, we will ask you a few short questions about your views related to the economy.

How will I get remunerated for my time? We highly value your participation in this 25-minute experiment, which is why we would like to remunerate you fairly. You will get £4 for participating in this survey. In addition, you have the chance to get up to £6 added to your payment. The more accurately you forecast, the more of the additional £6 you will take home. The maximum amount of money you could get is thus £10, and the minimum is £4. We will reward you as quickly as possible. Typically, you can expect to get rewarded within 1-3 business days after submission.

How will my data be used? Your answers will be completely anonymous, and we will keep them confidential. Your data will be stored in a password-protected file and may be used in academic publications. Your IP address will never be stored. Research data will be stored for a minimum of five years after publication or public release. By submitting your personal data, you agree to this transfer, storing or processing.

Who will have access to my data? Data are collected for the sole purpose of research and processed by researchers at the University of Oxford. Any personal information that could identify you will be removed or changed before files are shared with other researchers or results are made public. Responsible members of the University of Oxford and funders may be given access to data for monitoring and/or audit of the study to ensure we are complying with guidelines.

Do I have to take part? No. Please note that your participation is voluntary. You may withdraw at any point during the experiment for any reason, before submitting your answers, by closing the browser. However, we are only able to reimburse participants who complete the full survey.

Who has reviewed this project? This project has been reviewed by, and received ethics clearance through, the University of Oxford's Economics Department's Research Ethics Committee [reference number: ECONCIA21-22-24].

What if there is a problem? If you have a concern about any aspect of this project, please speak to Alena Wabitsch (alena.wabitsch@economics.ox.ac.uk), who will try to answer your query. The researcher should acknowledge your concern within 5 working days and give you an indication of how they intend to deal with it. If you remain unhappy or wish to make a formal complaint, please contact the relevant Chair of the Research Ethics Committee at the University of Oxford:

The Department of Economics Departmental Research Ethics Committee can be contacted at:

Email: ethics@economics.ox.ac.uk
Address: Department of Economics
Manor Road 10
Oxford OX1 3UQ
United Kingdom

The University of Oxford's Data Protection Officer can be contacted at:

Email address: data.protection@admin.ox.ac.uk
Telephone number: + 44 1865 280 199
Address: Planning and Council Secretariat
University Offices
Wellington Square
Oxford OX1 2JD
United Kingdom

If you have read and understood the information above, and agree to participate with the understanding that the data (including any personal data) you submit will be processed accordingly, please check the relevant box below to get started.

☐ Yes, I have understood the above information and agree to take part.

Please provide your Prolific ID:

Click Next to start the experiment.

Next

Figure A.14. Instructions (1)

Instructions

Please read these instructions carefully to make good decisions during the experiment. The quality of your decisions directly impact your bonus payment, which can be up to £6, in addition to the £4 participation payment. There will be a quiz on these instructions on the next page. If you answer at least one question incorrectly three times, the experiment will be ended early.

Throughout the experiment, you can always access these instructions by clicking the "Show Instructions" button at the bottom of any page.

Your Goal in the Experiment:

Your task is to **accurately forecast the annual inflation rate in the Euro area**. This means, you will be predicting the percentage increase or decrease in prices in the entire Euro area between a specific point in time and exactly 1 year earlier. Your goal is to **minimise forecast error**, which is the difference between your forecasted value of annual inflation and the realised annual inflation.

*Note: The more accurate your inflation forecasts, the lower your forecast error, the **more bonus payment** you get!*

Whenever we talk about inflation in this experiment, what we mean is the *annual* inflation rate (i.e., the percentage change in prices of goods and services over the course of exactly 1 year.) Data for inflation reflect real economic scenarios of the Euro area. Sometimes you will be shown professional forecasts, which are based on analyses of real European experts for the given, real economic scenarios.

The Experiment:

You will make two types of inflation forecasts:

- **Point Forecast:** Your 'Point Forecast' of inflation is your best guess of the exact value of inflation for the next unknown period.
- **Range Forecast:** Your 'Range Forecast' of inflation is the range of plausible values that you think will almost certainly contain the realised value of inflation in the next period (i.e. the upper and the lower inflation bounds of inflation).

The closer your point forecast is to the realised inflation, the more money you get. Similarly, the smaller your indicated range of inflation, the more money you get (as long as the realised inflation rate is inside your forecasted range).

What to Expect:

You will face **6 distinct** forecasting tasks. Each task is independent and unrelated to the others. In each task, you will first make your inflation forecasts ('*Initial Forecasts*'). You will then receive additional information about the next period's inflation rate and have a chance to adjust your forecasts ('*Updated Forecasts*'). After all 6 forecasting tasks, we will ask you a few questions, and then unveil the realised inflation rates and your performance.

What to do in each Forecasting Task:

1. Examine the given 10 periods of inflation history.
2. Make your Initial Forecasts for the next period:
 - A point forecast of inflation
 - The corresponding range forecast of inflation
3. You will be provided professional inflation forecasts, which are based on real historic forecasts of European experts for the respective scenarios. *Note: The provided historic differences between the realised inflation and the forecasted inflation represent the experts' historic forecast errors.*
4. Make your Updated Forecasts for the next period:
 - As before, provide a point forecast, and the corresponding range forecast.
 - *Note: Your Updated Forecasts can be the same as your Initial Forecasts, use some of the same values, or use completely different values.*
 - We will remind you of the values of your Initial Forecasts, when you are making your Updated Forecasts.

Overview of all the steps of a forecasting task:

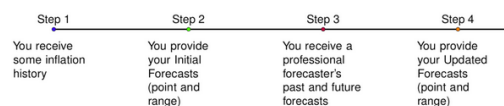


Figure A.15. Instructions (2)

Interacting with the data:

Historical data, professional forecasting history, and your point and range forecasts will be visualised in an interactive chart.

- Hover over chart dots for values.
- Type all your forecasts in percentages (up to 1 decimal) into the clearly labelled input text boxes.
- You may submit positive values (prices are going up), negative values (prices are going down), or a value of zero (prices are staying the same).
- Your forecast range can be as big or small as you prefer.
- Your point forecast doesn't have to be the mid-point of your range forecast. Your point forecast may be closer or farther from the upper than the lower bound.
- Your point forecast must always lie within your forecasted range - the software will prevent other inputs.
- Your point inflation forecasts will be visualised with green squares, and the corresponding range forecasts with green triangles.
- Submit your forecasts by clicking the 'Next' button. Note: You won't be able to change your answers after submitting them.

How your decisions get you bonus payment:

You can earn up to £6 in bonus payment for the accuracy of your point and range forecasts. The full £6 bonus requires your point forecast to match realised inflation (down to 1 decimal), and your range forecast to have the upper and lower bound on the value of the realised inflation (also down to 1 decimal).

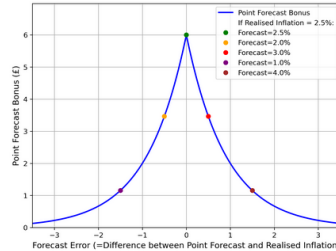
Your bonus will be awarded for **either** your point forecast accuracy **or** your range forecast accuracy in **one** set of forecasts (*Initial or Updated*) in **one of the 6** forecasting tasks. Here's how it works:

- Our software *randomly* selects one of your six forecasting tasks.
- For that forecasting task, the software *randomly* chooses either your Initial Forecasts or your Updated Forecasts.
- Within these selected Forecasts, the software *randomly* picks either your point forecast or your range forecast.

*Note: This means that to get a high bonus payment, you must take **both** the point and range forecasts in **all** Initial and Updated Forecasts across **all** forecasting tasks equally seriously. This is because you will only be paid for **one** randomly selected decision across the entire experiment.*

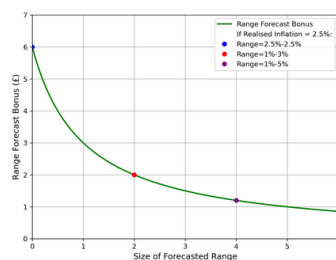
Point forecast bonus:

- Your exact forecast payment is computed like this: $P = 6 \cdot 3^{-|\text{Forecast Error}|}$
- The larger your forecast error (distance of forecasted inflation above or below the realised inflation), the less you earn.
- The figure below shows how your point forecast bonus falls, the further away your forecast is from the realised inflation value.
- The colourful dots in the figure visualise the following example: Suppose that realised inflation turns out to be **2.5%**:
 - If your point forecast of inflation was 2.5%, you would earn $P = 6 \cdot 3^{-0} = £6.00$
 - If your point forecast of inflation was 2.0%, you would earn $P = 6 \cdot 3^{-0.5} = £3.46$
 - If your point forecast of inflation was 3.0%, you would earn $P = 6 \cdot 3^{-0.5} = £3.46$
 - If your point forecast of inflation was 1.0%, you would earn $P = 6 \cdot 3^{-1.5} = £1.15$
 - If your point forecast of inflation was 4.0%, you would earn $P = 6 \cdot 3^{-1.5} = £1.15$



Range forecast bonus:

- If realised inflation is inside your range forecast, you score $R = 6 \cdot [1 / (1 + \text{total range})]$
- The total range of your forecast is given by the gap between the upper bound of the range forecast and the lower bound of the range forecast.
- The larger the range you create, the less money you earn, when inflation is within your range forecast.
- If inflation falls **outside** your forecast range, you earn **nothing**
- The figure below shows how your range forecast bonus falls, the larger your forecasted range is (as long as the realised value of inflation falls within that range).
- The colourful dots in the figure visualise the following example: Suppose that realised inflation turns out to be **2.5%**:
 - If your forecasted range was between 2.5% and 2.5%, then you would earn $R = 6 \cdot [1 / (1 + 0)] = £6.00$
 - If your forecasted range was between 1% and 3%, you would earn $R = 6 \cdot [1 / (1 + 2)] = £2.00$
 - If your forecasted range was between 1% and 5%, you would earn $R = 6 \cdot [1 / (1 + 4)] = £1.20$
 - If your forecasted range was between 0% and 2%, you would earn nothing because realised inflation is not within your range.



Note that the "Next" button may appear with a delay on certain pages. Similarly, parts of certain pages will be displayed only after a few seconds. Go through the experiment at your own pace and don't rush any decision.

Figure A.16. Comprehension Quiz

Comprehension Quiz

Before starting with the first forecasting tasks, we ask that you successfully complete the following comprehension quiz. You are always allowed to consult the instructions whenever you need to (see button at the bottom of the page).

If you get an answer wrong, you will be told so, and will have the opportunity to correct your answer. If you submit the quiz three times with at least one wrong answer, we will end the experiment early.

Your performance on this quiz does not affect your payment in any way, but you will not be able to proceed to the forecasting tasks until you have correctly answered all of the quiz questions.

1. True or False? Your goal in the experiment is to forecast inflation as accurately as possible.

☐ True ☐ False

2. True or False? For your bonus payment the software randomly selects a single forecasting task, of which it further randomly selects the Initial Forecasts or the Updated Forecasts, and finally it randomly selects either your point or your range forecast.

☐ True ☐ False

3. If inflation turns out to be 5.0% in the period selected for bonus payment, and your forecasted range is between -0.2% and 4.7%, how much extra payment will you get for your range?

4. True or False? If inflation turns out to be 2.2%, and your point forecast (that was selected for bonus payment) is 1.0%, your bonus payment is less than if you had forecasted 0.3%.

☐ True ☐ False

5. How many forecasting tasks will you have to do?

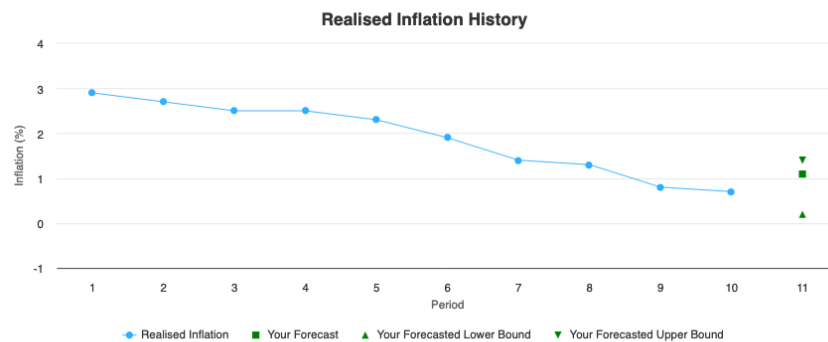
 

Next

Show instructions

Figure A.17. Prior Elicitation

Forecasting Task 1: Initial Forecasts



Provide your **point forecast** of inflation in period 11:

1.1 %

Provide the **highest probable value** of inflation in period 11 (i.e. the **upper bound of the range** you forecast inflation to fall in):

1.4 %

Provide the **lowest probable value** of inflation in period 11 (i.e. the **lower bound of the range** you forecast inflation to fall in):

0.2 %

Note: Enter any numeric value. You may use up to 1 decimal. Indicate decimals with a dot ("."). Use the "-" sign for negative values. The chart will automatically update its scaling to visualise your inputs.

Next

Show instructions

Notes: Participants are forced to spend at least 10 seconds on this page to reduce the risk of rushing decisions without paying attention to the provided information.

Figure A.18. Posterior Elicitation

Forecasting Task 1: Updated Forecasts

Now imagine an expert from Italy who represents the European Central Bank (ECB) provides a forecast of 1.2% for inflation in period 11.

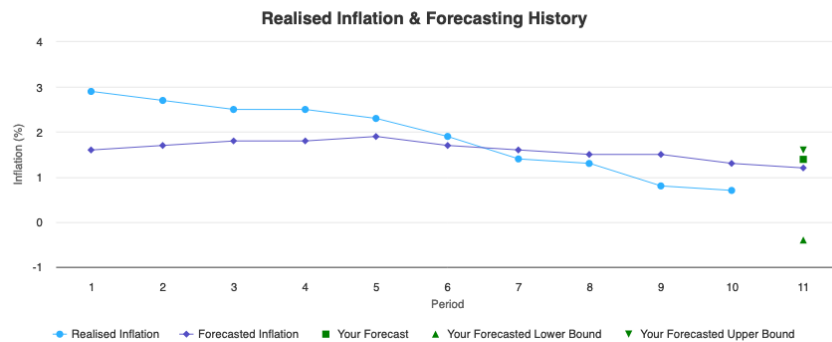
You find this forecast, as well as the expert's corresponding forecasting history, displayed in the graph.

In addition, you may click on any or all of the 3 "Read more" buttons below for detailed explanations by the expert from Italy who represents the European Central Bank (ECB) about the analyses underlying the forecast for period 11.

Read more

Read more

Read more



You forecasted 1.1% inflation for period 11. You may now keep or revise this point forecast:

1.4 %

You forecasted the **highest probable value** of inflation in period 11 to be 1.4%. You may now keep or revise this **upper bound** of your range forecast:

1.6 %

You forecasted the **lowest probable value** of inflation in period 11 to be 0.2%. You may now keep or revise this **lower bound** of your range forecast:

-0.4 %

Note: Enter any numeric value. You may use up to 1 decimal. Indicate decimals with a dot ("."). Use the "-" sign for negative values. The chart will automatically update its scaling to visualise your inputs.

Next

Show instructions

Notes: The first sentence of this page is shown in isolation for 5 seconds before the rest of the page appears to increase reporting on the messenger and signal. Participants are then forced to spend at least another 10 seconds on this page to reduce the risk of rushing decisions without paying attention to the provided information. Hovering over the dots in the graph reveals exact values of realized inflation and inflation forecasts.

Figure A.19. Attention Check

Attention Check

The following questions assess your attentiveness during the **last (most recent)** of the 6 forecasting tasks you just completed. Please answer to the best of your ability. Your responses will not impact your final payment.

Which expert provided the forecast in the last forecasting task?

What value of inflation for Period 11 did the expert forecast in the last forecasting task?

Next

Show instructions

Notes: The order of answer options for experts is randomized. The options for period 11 values reflect actual signal values of the 6 forecasting tasks.

Figure A.20. Survey: Purpose of Experiment & Strategy

Survey

Before we reveal the correct values for inflation and the bonus payment that you have earned, we would like to ask you a few questions.

We are interested in your honest opinions and perceptions. Therefore, there are no right or wrong answers.

Note: You may leave any of the comment boxes below empty, but we highly appreciate hearing your opinion.

What do you think is the purpose of this study?

What - if any - strategy did you apply to your decision-making to maximise the bonus payment from this study?

Next

Show instructions

Notes: Participants are allowed to skip any or all of these questions entirely.

Figure A.21. Survey: Association

Survey

What - if anything - did you **think about**, or associate with, the **experts** who provided you the forecasts?

Write down **anything that comes to mind**, in any form you would like. We are also interested to hear if you did not think about an expert at all.

Keywords are sufficient - you do not have to write full sentences. You may leave some boxes empty, but we highly appreciate to hear about your views and perceptions.

Note: There are no right or wrong answers.

The expert from the European Central Bank (ECB):

The Spanish expert from the European Central Bank (ECB):

The French expert:

The Spanish expert:

The French expert from the European Central Bank (ECB):

The expert from the Banque de France:

Next

Show instructions

Notes: Participants are allowed to skip any or all of these questions entirely. To make it easier for participants to recall, the order of experts represents the occurrence of experts in forecasting tasks (i.e., randomized across individuals).

Figure A.22. Survey: Ability

Survey

How confident were you in each **expert's ability** to accurately predict inflation and provide accurate economic analysis?

Give your answer **on a scale of 1 to 7**, where **1** represents **Complete Lack in Confidence** in their abilities, **4** represents being **Neutral** about their abilities and **7** represents **Complete Confidence** in their abilities.

Note: Again, there are no right or wrong answers.

The expert from the European Central Bank (ECB):

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

The Spanish expert from the European Central Bank (ECB):

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

The French expert:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

The Spanish expert:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

The French expert from the European Central Bank (ECB):

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

The expert from the Banque de France:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

Next

Show instructions

Notes: Order of experts represents occurrence of experts in forecasting tasks (i.e., randomized across individuals).

Figure A.23. Survey: Exposure

Survey

How well are you **informed** about the **institutions** or **policymakers** listed below? Give your answer **on a scale of 0 to 5**, where these values represent the following:

1: I have **never** heard of them before this experiment.

2: I have heard of them, but I **don't know with** certainty what they do.

3: I **know** with certainty what/who they are and what they do, but I **don't get any news** about them.

4: I **know** with certainty what/who they are and what they do, and I get news **occasionally** about them.

5: I **know** with certainty what/who they are and what they do, and I get news **regularly** about them.

Remember: There are no right or wrong answers.

Pablo Hernández de Cos:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Joachim Nagel:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

The Banque de France (the national central bank of France):

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

François Villeroy de Galhau:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Fabio Panetta:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Ignazio Visco:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Luis de Guindos:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Isabel Schnabel:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Christine Lagarde:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

The European Central Bank (ECB):

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Next

Show instructions

Notes: The order in which institutions and policymakers appear is randomized at the individual-level.

Figure A.24. Survey: Trust

Survey

How much do you **trust** the institutions or policymakers listed below?

Give your answer **on a scale of 1 to 7**, where **1** represents **Completely Distrusting** them, **4** represents being **Neutral** (neither trusting nor distrusting them) and **7** represents **Completely Trusting** them.

Note: Again, there are no right or wrong answers.

The Banque de France (the national central bank of France):

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

Isabel Schnabel:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

Christine Lagarde:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

Luis de Guindos:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

Ignazio Visco:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

Joachim Nagel:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

François Villeroy de Galhau:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

Fabio Panetta:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

The European Central Bank (ECB):

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

Pablo Hernández de Cos:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

[Next](#)

[Show instructions](#)

Notes: This page is only shown to participants who indicate knowing at least one institution of policymaker. Only institutions and policymakers that the participant indicates knowing appear in this list, as indicated trust is only meaningful if participants know the institutions and policymakers. The order in which institutions and policymakers appear is randomized at the individual-level.

Figure A.25. Survey: Monetary Policy Expertise

Survey

This is our last question. On the next page, we will show you how well you did in the forecasting tasks, and thus how much money you made in this experiment.

Note: Please select a single answer to the question below. Again, there are no right or wrong answers.

Every 6 weeks, the Governing Council of the ECB makes a monetary policy decision for the Euro area, which may include potential changes to key interest rates. Subsequently, the ECB announces this decision during a press conference.

On average, over the past 8 years, how often have you been aware of these monetary policy decisions by the ECB (even if there was no change in interest rates)? This can be through receiving news in any form of media, from the ECB directly, through discussions with others, or any other information sources.

- ☐ every 6 weeks (ca. at least 8-9 times a year)
- ☐ every 6-12 weeks (ca. 5-7 times a year)
- ☐ every 12-18 weeks (ca. 3-4 times a year)
- ☐ every 18-24 weeks (ca. once or twice a year)
- ☐ less than once a year

Next

Show instructions

Notes: The question is allows for potential comparison of experimental and Twitter results, distinguishing between experts and non-experts. With some caveats, answers to this questions can further be seen as an imperfect proxy for monetary policy expertise.

Figure A.26. Payoff Overview

Realised Inflation & Your Performance

We now reveal the realised values for inflation in period 11, your corresponding forecasts, and the bonus payment that you will earn, if the software selects a forecast.

Forecasting Task 1: The realised inflation was 2.1%.

	Initial Forecasts		Updated Forecasts	
	Your Forecasts	Bonus Payments	Your Forecasts	Bonus Payments
Point Forecasts	1.0%	£1.79	1.0%	£1.79
Range Forecasts	-0.2%-1.4%	£0.0	1.0%-2.0%	£0.0

Forecasting Task 2: The realised inflation was 2.2%.

	Initial Forecasts		Updated Forecasts	
	Your Forecasts	Bonus Payments	Your Forecasts	Bonus Payments
Point Forecasts	1.1%	£1.79	1.0%	£1.61
Range Forecasts	1.1%-2.9%	£2.14	-124.0%-1.0%	£0.0

Forecasting Task 3: The realised inflation was 1.0%.

	Initial Forecasts		Updated Forecasts	
	Your Forecasts	Bonus Payments	Your Forecasts	Bonus Payments
Point Forecasts	1.1%	£5.38	1.0%	£6.0
Range Forecasts	0.0%-2.9%	£1.54	1.0%-1.0%	£6.0

Forecasting Task 4: The realised inflation was 1.1%.

	Initial Forecasts		Updated Forecasts	
	Your Forecasts	Bonus Payments	Your Forecasts	Bonus Payments
Point Forecasts	1.1%	£6.0	1.0%	£5.38
Range Forecasts	-0.2%-2.2%	£1.76	1.0%-1.0%	£0.0

Forecasting Task 5: The realised inflation was 1.7%.

	Initial Forecasts		Updated Forecasts	
	Your Forecasts	Bonus Payments	Your Forecasts	Bonus Payments
Point Forecasts	1.6%	£5.38	1.6%	£5.38
Range Forecasts	0.0%-2.9%	£1.54	0.4%-2.5%	£1.94

Forecasting Task 6: The realised inflation was 0.6%.

	Initial Forecasts		Updated Forecasts	
	Your Forecasts	Bonus Payments	Your Forecasts	Bonus Payments
Point Forecasts	1.0%	£3.87	1.0%	£3.87
Range Forecasts	1.0%-1.0%	£0.0	1.0%-1.0%	£0.0

Click the "Next" button below for the software to randomly select the forecast for your bonus payment.

Next

Show instructions

Notes: Realized inflation values for period 11 and the corresponding performance of participants are only shown at the very end of the experiment and after all survey questions to avoid learning effects between forecasting tasks and performance-driven biases to survey answers asked after the experiment.

Figure A.27. Payoff Reveal

Your payment

The software randomly selected your **Range Forecast** of the **Updated Forecast of Forecasting Task 1** for your bonus payment.

Given your performance for this Range forecast, your bonus payment will thus be £0.00.

Together with the participation payment of £4, you will be paid a **total of £4.00** for your participation in this experiment.

Please let us know of any comment you may have in the text box below, and click the "Next" button below to end the experiment.

Next

Show instructions

Notes: The software reveals the randomly selected a decision for each participant's bonus payment.

A.4 Model Details

A.4.1 Threshold for Public Signal Availability

To model the variability in agents' thresholds, I use the tractable and convenient truncated normal distribution. Each outgroup agent's threshold d_o follows a truncated normal distribution $\mathcal{N}_+(0, 1)$ on $[0, \infty)$, which ensures thresholds are non-negative. The probability density function (PDF) of d_o is:

$$f_{d_o}(d) = \frac{\phi(d)}{\Phi(\infty) - \Phi(0)} = 2\phi(d), \quad \text{for } d \geq 0,$$

where $\phi(d)$ and $\Phi(d)$ are the standard normal PDF and cumulative distribution function (CDF), respectively. The cumulative distribution function (CDF) of d_o is:

$$F_{d_o}(d) = \int_0^d f_{d_o}(z) dz = 2\Phi(d) - 1.$$

This CDF determines the fraction of outgroup agents who observe Y based on their individual thresholds.

A.4.2 Proofs and Derivations

A.4.2.1 Proof of Lemma 1: Optimal Signal Extraction Weights

Lemma 1 claims that the optimal weights κ_h for ingroup and informed outgroup agents are given by:

$$\kappa_g^* = \frac{\tau_y q}{\tau_Y + \tau_y q} \quad \text{and} \quad \kappa_o^* = \frac{\tau_y q}{\rho_{im} \tau_Y + \tau_y q},$$

where $q = (1 - r + r(1 - \alpha)(1 - A))$.

Below outlines the corresponding proof of these optimal signal extraction weights.

Uninformed Outgroup Agents An uninformed outgroup agent simply chooses $a_{io}(y_i) = y_i$. Since Y is unobserved by this agent, he relies on $E[Y] = x$, so he uses $E_{io}[x | y_i] = y_i$ for any Y .

To derive the action of informed agents, suppose that the population, consisting of the different agent types h , follows a linear strategy of the form:

$$a_{jh} = \kappa_h y_j + (1 - \kappa_h) Y. \tag{A.7}$$

Ingroup Agents An ingroup agent g , who always receives the public signal, updates without bias and expects all agents in the economy to update fully Bayesian. She assumes that both ingroup agents and informed outgroup agents who receive the public signal Y use the same signal extraction weight κ_g . While she knows that uninformed outgroup agents rely solely on their private signals, she believes that informed outgroup agents process the public signal in the same unbiased way as she does. In other words, the ingroup agent is unaware of any biases or differences in information processing among other agents and expects all agents to update their beliefs and choose actions just like herself, given their information sets. Knowing the fraction of outgroup agents A that receive the public signal and the overall share of outgroup agents $(1 - \alpha)$, her expected average action of the population is:

$$\begin{aligned}
E_{ig}[\bar{a} \mid y_i, Y] &= [\alpha\kappa_g + (1 - \alpha)A\kappa_g + (1 - \alpha)(1 - A)] E_{ig}[y_j \mid y_i, Y] \\
&\quad + [\alpha(1 - \kappa_g) + (1 - \alpha)A(1 - \kappa_g)] Y \\
&= [\alpha\kappa_g + (1 - \alpha)A\kappa_g + (1 - \alpha)(1 - A)] \frac{\tau_y y_i + \tau_Y Y}{\tau_y + \tau_Y} \\
&\quad + [\alpha(1 - \kappa_g) + (1 - \alpha)A(1 - \kappa_g)] Y \\
&= Y \left[(1 - \kappa_g) [\alpha + (1 - \alpha)A] + \frac{\tau_Y}{\tau_y + \tau_Y} [\kappa_g (\alpha + (1 - \alpha)A) + (1 - \alpha)(1 - A)] \right] \\
&\quad + y_i \left[\frac{\tau_y}{\tau_y + \tau_Y} [\kappa_g (\alpha + (1 - \alpha)A) + (1 - \alpha)(1 - A)] \right].
\end{aligned} \tag{A.8}$$

Thus, the ingroup agent g 's optimal action is given by:

$$\begin{aligned}
a_{ig}(y_i, Y) &= (1 - r)E_{ig}[x \mid y_i, Y] + rE_{ig}[\bar{a} \mid y_i, Y] \\
&= r \left\{ Y \left[(1 - \kappa_g) (\alpha + (1 - \alpha)A) + \frac{\tau_Y}{\tau_y + \tau_Y} [\kappa_g (\alpha + (1 - \alpha)A) + (1 - \alpha)(1 - A)] \right] \right. \\
&\quad \left. + y_i \left[\frac{\tau_y}{\tau_y + \tau_Y} [\kappa_g (\alpha + (1 - \alpha)A) + (1 - \alpha)(1 - A)] \right] \right\} + (1 - r) \frac{\tau_y y_i + \tau_Y Y}{\tau_y + \tau_Y} \\
&= \left[\frac{\tau_y}{\tau_y + \tau_Y} (1 - r + r[\kappa_g (\alpha + (1 - \alpha)A) + (1 - \alpha)(1 - A)]) \right] y_i \\
&\quad + \left[1 + \frac{\tau_y}{\tau_y + \tau_Y} [r[-\kappa_g (\alpha + (1 - \alpha)A) + \alpha + (1 - \alpha)A] - 1] \right] Y.
\end{aligned} \tag{A.9}$$

Comparing coefficients with the population's strategy (Equation (A.7)) gives:

$$\kappa_g = \frac{\tau_y}{\tau_y + \tau_Y} [1 - r + r [\kappa_g (\alpha + (1 - \alpha)A) + (1 - \alpha)(1 - A)]] , \tag{A.10}$$

from which I can solve for κ_g :

$$\kappa_g^* = \frac{\tau_y q}{\tau_Y + \tau_y q}. \tag{A.11}$$

Informed Outgroup Agents The informed outgroup agent o —who receives both signals but updates with a bias—looks very similar to the ingroup agent but down-weights the public signal by the resonance weight ρ_{im} . His expected average action of the population is:

$$\begin{aligned}
E_{io}[\bar{a} \mid y_i, Y] &= [\alpha\kappa_o + (1 - \alpha)A\kappa_o + (1 - \alpha)(1 - A)] E_{io}[y_j \mid y_i, Y] \\
&\quad + [\alpha(1 - \kappa_o) + (1 - \alpha)A(1 - \kappa_o)] Y \\
&= \alpha[\kappa_o E_{io}[y_j \mid y_i, Y] + (1 - \kappa_o)Y] + (1 - \alpha)A[\kappa_o E_{io}[y_j \mid y_i, Y] + (1 - \kappa_o)Y] \\
&\quad + (1 - \alpha)(1 - A)E_{io}[y_j \mid y_i, Y] \\
&= Y \left[(1 - \kappa_o) [\alpha + (1 - \alpha)A] + \frac{\rho_{im}\tau_Y}{\tau_y + \rho_{im}\tau_Y} [\kappa_o (\alpha + (1 - \alpha)A) + (1 - \alpha)(1 - A)] \right] \\
&\quad + y_i \left[\frac{\tau_y}{\tau_y + \rho_{im}\tau_Y} [\kappa_o (\alpha + (1 - \alpha)A) + (1 - \alpha)(1 - A)] \right].
\end{aligned} \tag{A.12}$$

Thus, the informed outgroup agent o 's optimal action is given by:

$$\begin{aligned}
a_{io}(y_i, Y) &= (1 - r)E_{io}[x \mid y_i, Y] + rE_{io}[\bar{a} \mid y_i, Y] \\
&= \left[\frac{\tau_y}{\tau_y + \rho_{im}\tau_Y} (1 - r + r[\kappa_o (\alpha + (1 - \alpha)A) + (1 - \alpha)(1 - A)]) \right] y_i \\
&\quad + \left[1 + \frac{\tau_y}{\tau_y + \rho_{im}\tau_Y} [r[-\kappa_o (\alpha + (1 - \alpha)A) + \alpha + (1 - \alpha)A] - 1] \right] Y.
\end{aligned} \tag{A.13}$$

Comparing coefficients with the population's strategy (Equation (A.7)), I get for the informed outgroup agent:

$$\kappa_o = \frac{\tau_y}{\tau_y + \rho_{im}\tau_Y} [1 - r + r[\kappa_o (\alpha + (1 - \alpha)A) + (1 - \alpha)(1 - A)]], \tag{A.14}$$

from which I can solve for κ_o :

$$\kappa_o^* = \frac{\tau_y q}{\rho_{im}\tau_Y + \tau_y q}. \tag{A.15}$$

Therefore, the optimal weights κ_h for ingroup and informed outgroup agents are as stated in Lemma 1.

A.4.2.2 Proof of Proposition 1: Uniqueness of Linear Equilibrium

Proposition 1 claims the equilibrium strategies derived constitute the unique linear equilibrium of the game.

Proof: To prove that the equilibrium strategy is indeed a linear combination of the available signals, and that the equilibrium is therefore unique, I extend the proof by Morris and Shin (2002) to accommodate heterogeneous agent types with biased information processing. The

optimal action of agents is given by Equation (1.13), and the average action is given in Equation (1.21), which can also be written as:

$$\begin{aligned}\bar{a} &= \int_0^\alpha a_{ig}(y_i, Y) di + A \int_\alpha^1 a_{io}(y_i, Y) di + (1 - A) \int_\alpha^1 a_{io}(y_i) di \\ &= \alpha \bar{a}_{ig}(x, Y) + (1 - \alpha) A \bar{a}_{io}(x, Y) + (1 - \alpha)(1 - A)x.\end{aligned}\tag{A.16}$$

Integrating over the interval $[0, \alpha]$ for ingroup agents and $[\alpha, 1]$ for outgroup agents returns the average action as a weighted sum of the expected actions of each group, given the symmetry and independence of agents within each group. The integrals over outgroup agents are weighted by A and $(1 - A)$, which represent the fractions of informed and uninformed agents, respectively. Remember that agents expect other informed agents to update and act just like themselves, as agents are naive about their own biases and those of others. This means that ingroup agents assume the average action of informed agents is $\bar{a}_i(x, Y) = \bar{a}_{ig}(x, Y)$, while informed outgroup agents assume $\bar{a}_i(x, Y) = \bar{a}_{io}(x, Y)$ across all informed agents. The average action of uninformed outgroup agents is known by all and is $\bar{a}_{io}(x) = x$, as defined earlier. Hence, the optimal strategy of any informed agent of type h is:

$$\begin{aligned}a_{ih}(y_i, Y) &= qE_{ih}[x|y_i, Y] + [r(\alpha + (1 - \alpha)A)]E_{ih}[\bar{a}|y_i, Y] \\ &= qE_{ih}[x|y_i, Y] + [r(\alpha + (1 - \alpha)A)]qE_{ih}[\bar{E}(x)] + [r(\alpha + (1 - \alpha)A)]^2E_{ih}[\bar{E}[\bar{a}(x, Y)]], \quad j \neq i \\ &= q \sum_{s=0}^{\infty} (r(\alpha + (1 - \alpha)A))^s E_{ih}[\bar{E}_i^s[x|y_i, Y]].\end{aligned}\tag{A.17}$$

$\bar{E}_i^s[x]$ denotes the average expectation of the average expectation (s -times) of x , and $\bar{E}_i^s[\bar{a}]$ is the average expectation of the average expectation (s -times) of the average action, where all averages concern informed agents only and are dependent on agent type h . The infinite sum captures the recursiveness of higher-order beliefs, where agents form expectations about the expectations of others in this strategic interaction setting. $E_{ih}[\bar{E}_i^s[x|y_i, Y]]$ is a linear combination of y_i and Y .¹¹ Thus, $a_{ih}(y_i, Y)$ is also a linear combination of these signals. By establishing that the optimal actions of agents are linear combinations of their signals, and that these strategies lead to consistent expectations about others' actions, it is shown that the linear solution constitutes the unique equilibrium.

A.4.2.3 Private Signal Precision May Harm Welfare

Proposition 4. *The effect of increasing the precision of private signals (τ_y) on expected welfare is ambiguous and depends on the coordination parameter r and the presence of outgroup agents ($\frac{\partial E(W|x)}{\partial \tau_y} \gtrless 0$).*

¹¹The proof by Morris and Shin (2002) is directly applicable to ingroup agents. It extends to expectation formation with the resonance weight ρ_{im} of informed outgroup agents as the recursive structure and the linearity of the expectation function (Equation (1.9)) are preserved. The updating bias merely reduces the influence of the public signal, and thus shifts the coefficients in the explicit linear expression slightly.

Derivation: Taking the derivative of $E[W(a, x)|x]$ with respect to τ_y shows that $\frac{\partial E[W|x]}{\partial \tau_y} \geq 0$ and depends on the particular parameter values.

$$\begin{aligned} \frac{\partial E(W | x)}{\partial \tau_y} = & -\frac{\alpha q^2}{(\tau_Y + \tau_y q)^2} - \frac{(1 - \alpha) A q^2}{(\rho_{im} \tau_Y + \tau_y q)^2} \\ & + \frac{2 A q (1 - \alpha) (\tau_y q^2 + \rho_{im}^2 \tau_Y)}{(\rho_{im} \tau_Y + \tau_y q)^3} \\ & + \frac{2 q \alpha (\tau_y q)^2 + \tau_Y}{(\tau_Y + \tau_y q)^3} + \frac{(1 - \alpha)(1 - A)}{\tau_y^2} \end{aligned} \quad (A.18)$$

At first glance, more precise private signals should enhance welfare by improving alignment with x . However, due to biased information processing and strategic complementarities, increasing τ_y can reduce welfare. Outgroup agents, unaware of their bias, overestimate others' reliance on private signals and overemphasize their own private information to coordinate. This leads to misalignment with the fundamental and increased action dispersion, especially when strategic complementarities are strong and the outgroup's updating bias is severe. The resonance weight required for negative welfare effects to materialize is much lower than what I find empirically.¹²

This finding contrasts with Morris and Shin (2002) and even Cornand and Heinemann (2008), who also restrict public signal availability, but aligns with other work such as Hellwig (2005) and Angeletos and Pavan (2007), finding that better private information increases reliance on individual signals, worsening coordination and increasing action dispersion. Specifically, Hellwig (2005) introduces heterogeneous private information, and Angeletos and Pavan (2007) show that welfare effects depend on the relationship between equilibrium and efficient coordination degrees.

Corollary 2. *When there are significant biases in outgroup agents' information processing, increasing τ_y may reduce social welfare.*

A.4.2.4 Derivation for Proposition 2: Welfare Implications of the Public Signal Precision

Taking the derivative of $E[W(a, x)|x]$ with respect to τ_Y shows that $\frac{\partial E[W|x]}{\partial \tau_Y} \geq 0$ and depends on the particular parameter values.

$$\begin{aligned} \frac{\partial E(W | x)}{\partial \tau_Y} = & \frac{2\alpha (\tau_Y + \tau_y q^2)}{(\tau_Y + \tau_y q)^3} - \frac{\alpha}{(\tau_Y + \tau_y q)^2} \\ & + \frac{2A(1 - \alpha)\rho_{im} (\rho_{im}^2 \tau_Y + \tau_y q^2)}{(\rho_{im} \tau_Y + \tau_y q)^3} - \frac{A(1 - \alpha)\rho_{im}^2}{(\rho_{im} \tau_Y + \tau_y q)^2} \end{aligned} \quad (A.19)$$

¹²Specifically, $\rho_{im} \lesssim 0.3$ is needed. This occurs for $r \gtrsim 0.5$, equally precise public and private information, and ensuring at least half of agents are outgroup agents who are almost surely informed. For other combinations negative welfare effects are even less of a concern.

A.4.2.5 Derivation for Proposition 3: The Optimal Share of Ingroup Agents

The optimal α^* is found by setting $\frac{\partial E[W|x]}{\partial \alpha} = 0$ and solving for α . Complex non-linearities makes an explicit analytical solution for α^* unfeasible. The Weierstrass Extreme Value Theorem guarantees the existence of a welfare-maximizing α^* , given that the expected welfare function $E[W(a, x)|x]$ is continuous in α and the domain $[0, 1]$ is compact. α^* is either a corner solution (0 or 1) or is obtained by setting the first-order condition to zero:

$$\begin{aligned} \frac{\partial E(W | x)}{\partial \alpha} = & \frac{\alpha 2r\tau_y (1-A)q}{(\tau_Y + \tau_y q)^2} - \frac{\alpha 2r\tau_y (1-A)(\tau_Y + \tau_y q^2)}{(\tau_Y + \tau_y q)^3} + \frac{(1-\alpha)2A(1-A)r\tau_y q}{(\rho_{im}\tau_Y + \tau_y q)^2} \\ & - \frac{(1-\alpha)2A(1-A)r\tau_y(\tau_y q^2 + \rho_{im}^2\tau_Y)}{(\rho_{im}\tau_Y + \tau_y q)^3} + \frac{A(\tau_y q^2 + \rho_{im}^2\tau_Y)}{(\rho_{im}\tau_Y + \tau_y q)^2} - \frac{\tau_Y + \tau_y q^2}{(\tau_Y + \tau_y q)^2} + \frac{1-A}{\tau_y} = 0. \end{aligned}$$

A.4.3 Details on Expertise Loss

Proposition 5. *Delegating communication to increase α can be welfare-improving if the marginal benefit outweighs the marginal cost from a reduction in public signal precision due to the potential lower expertise of the delegated messengers: $\frac{\partial E[W|x]}{\partial \alpha} > k \frac{\partial E[W|x]}{\partial \tau_Y}$.*

For delegation to be welfare-improving, the marginal benefit must outweigh the marginal cost: $\frac{\partial E[W|x]}{\partial \alpha} > k \frac{\partial E[W|x]}{\partial \tau_Y}$, where k represents the rate at which precision decreases as α increases. This occurs when either the direct improvement in agents' alignment with the fundamental (due to more ingroup agents) is significant, or when the loss in public signal precision (τ_Y) per unit increase in α (represented by k) is small, minimizing the marginal cost.

A.4.3.1 Proof Sketch of Proposition 5: Delegating Communication vs. Loss in Expertise

The trade-off is formalized by comparing the marginal benefit of increasing α with the marginal cost of reduced τ_Y : Suppose that increasing α (the share of ingroup agents) comes at the cost of reducing τ_Y (the precision of the public signal). This is because delegating communication to messengers who better match the characteristics of more agents might involve using messengers who are less expert or less effective communicators, thereby reducing the quality (precision) of the public signal.

Assume that τ_Y is a decreasing function of α :

$$\tau_Y = \tau_Y(\alpha), \quad \text{with} \quad \frac{d\tau_Y}{d\alpha} \leq 0.$$

Recall that expected social welfare is given by:

$$E[W(a, x)|x] = -\alpha \frac{\tau_Y + \tau_y q^2}{[\tau_Y + \tau_y q]^2} - (1-\alpha)A \frac{\rho_{im}^2 \tau_Y + \tau_y q^2}{[\rho_{im}\tau_Y + \tau_y q]^2} - \frac{(1-\alpha)(1-A)}{\tau_y}.$$

To find the condition under which increasing α leads to an increase in expected welfare—considering the dependence of τ_Y on α —I take the total derivative of expected welfare with respect to α is:

$$\frac{dE[W|x]}{d\alpha} = \frac{\partial E[W|x]}{\partial \alpha} + \frac{\partial E[W|x]}{\partial \tau_Y} \cdot \frac{d\tau_Y}{d\alpha}.$$

The goal is to determine whether $\frac{dE[W|x]}{d\alpha} \leq 0$ (welfare increases as α increases, since welfare is defined as a negative value of losses).

Equation (A.4.2.5) shows the expression for $\frac{\partial E[W|x]}{\partial \alpha}$. This partial derivative captures the direct effect of changing α on expected welfare, holding τ_Y constant, and represents the net change in welfare due to reallocating a marginal agent from the outgroup to the ingroup.

The expression for $\frac{\partial E[W|x]}{\partial \tau_Y}$ is shown in Equation (A.19). This partial derivative captures how changes in the precision of the public signal affect expected welfare, holding α constant.

To compute $\frac{d\tau_Y}{d\alpha}$, assume that τ_Y decreases linearly with α :

$$\tau_Y = \tau_Y^0 - k\alpha, \quad \text{with } k > 0 \quad \text{and} \quad 0 \leq \alpha \leq 1,$$

where τ_Y^0 is the initial precision of the public signal when $\alpha = 0$, and k represents the rate at which precision decreases as α increases. Then:

$$\frac{d\tau_Y}{d\alpha} = -k < 0.$$

Now, the total derivative is:

$$\frac{dE[W|x]}{d\alpha} = \frac{\partial E[W|x]}{\partial \alpha} - k \frac{\partial E[W|x]}{\partial \tau_Y}.$$

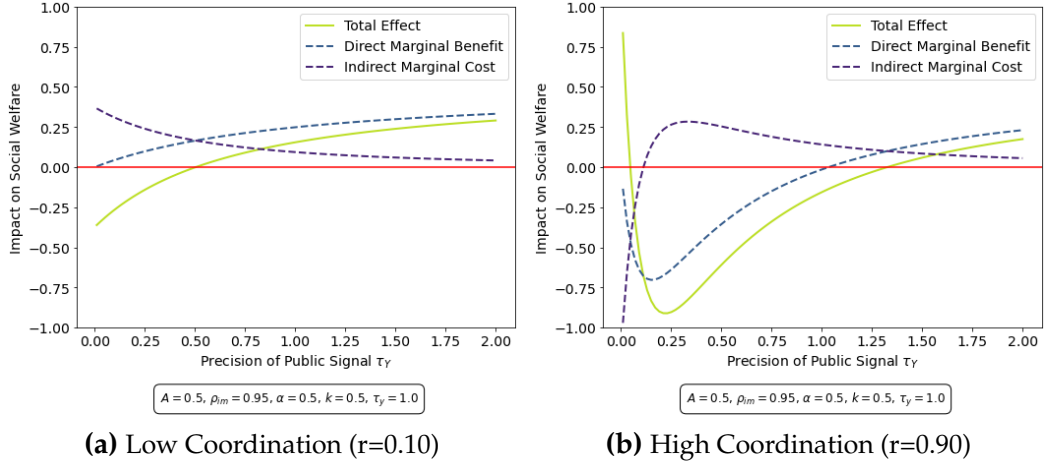
Under which conditions is the total derivative negative (since welfare is negative, a negative derivative means welfare increases)? The total effect of increasing α on welfare depends on:

1. *Direct Marginal Benefit* ($\frac{\partial E[W|x]}{\partial \alpha}$): Increasing α directly reduces the distance from the fundamental for ingroup agents $E[(\bar{a}_{ig}(x, Y) - x)^2]$ because these agents process the public signal without bias. It also reduces the weight on the higher distance from the fundamental of outgroup agents.
2. *Indirect Marginal Cost* ($-k \frac{\partial E[W|x]}{\partial \tau_Y}$): Increasing α reduces τ_Y , lowering the precision of the public signal. A lower τ_Y increases the distance from the fundamental $E[(\bar{a}_{ih}(x, Y) - x)^2]$ for both ingroup and any remaining informed outgroup agents.

The marginal benefit must outweigh the marginal cost for delegation to improve welfare:

$$\frac{\partial E[W|x]}{\partial \alpha} > k \frac{\partial E[W|x]}{\partial \tau_Y}.$$

Figure A.28. Social Welfare Impact of Delegation with Expertise Loss



Notes: Plots show impact on welfare by raising the share of ingroup agents via delegation, accounting for potential losses in the precision of the public signal due to lower expertise of the messenger. Green line shows the total impact, blue and purple dashed lines show the marginal benefit and marginal cost of delegating, respectively. τ_Y is set to 1, A , k and initial α are set to 0.5 each. Arrows indicate the direction in which welfare is increasing. Panel (a) shows a case of insufficient strategic complementarity ($r=0.10$). Panel (b) shows a case of sufficient strategic complementarity ($r=0.90$).

This condition formalizes the trade-off between the gain from enhanced coordination (marginal benefit) and the potential loss in public signal precision (marginal cost). If this inequality holds, increasing α improves expected welfare, proving Proposition 5. The specific conditions depend on the model parameters ($r, \tau_Y, \tau_Y^0, k, \rho_{im}, A$).

If $\frac{\partial E[W|x]}{\partial \alpha} > 0$ and $-k \frac{\partial E[W|x]}{\partial \tau_Y} < 0$, then the total derivative $\frac{dE[W|x]}{d\alpha} \geq 0$. Delegation is welfare-improving only when the marginal benefit of increasing α exceeds the marginal cost from reducing τ_Y . This either occurs when the direct improvement in agents' alignment with the fundamental is substantial, or when the marginal cost, which is the loss in public signal precision (τ_Y) per unit increase in α (represented by k), is sufficiently small.

Therefore, delegating communication to increase α is welfare-improving if the gain from enhanced coordination outweighs the potential loss in public signal precision due to less expertise of the delegated messengers.

Figure A.28 visualizes the trade-off between direct marginal benefits and indirect marginal costs of delegating to messengers with lower (perceived) expertise in very low and very high coordination environments.

Appendix to Chapter 2: Central Bank Communication With Non-Experts—A Road to Nowhere?

B.1 Details on the Data Construction

B.1.1 Cleaning Procedure for ECB-Related Tweets

To ensure tweets are indeed related to the European Central Bank, we start by looking at random subsamples of tweets and manually identify those that are unrelated. This provides a broad idea of what types of other tweets our data collection method extracted. With these unrelated tweets, certain words or phrases are established that distinguish them from observations that are indeed talking about central banking (for instance, the term “cricket” helps distinguishing tweets about the English Cricket Board from those about the European Central Bank, both of which are often abbreviated as ECB). Furthermore, a visual check is implemented using word clouds. Two clouds are created, one based on the cleaned sample and the other on dropped observations. The former cloud helps to check whether the majority of words is related to central banking, and it helps to identify other unrelated and frequent topics (such as cricket). The latter word cloud serves to check whether central banking-related tweets are not falsely excluded, and further, singles out further key words for the cleaning procedure (e.g. the names of cricket players) by displaying words that appear frequently in the unrelated set of tweets. Examples of such word clouds are found in Appendix Figure B.1. These steps are repeated until we are satisfied with the content of the final sample.

All tweets that contain the identified text in their body or hashtags are dropped. For instance, this removes tweets related to the English Cricket Board, as all tweets that contain certain names of cricket players, and terms like “cricket”, “skipper”, “sport”, “coach”, “batsman” or “ecb.co.uk” are filtered out. Further, tweets about the Extra Care Buck by the American drugstore chain CVS, a Samsung charger called “ECB-DU4EWE”, a camera case called

Panel A: English sample

[illegible][illegible]

“ECB-1 EVA”, a part of SharePoint (a Microsoft’s document management tool) called Edit Control Block and others are removed.

Next is a check on whether tweets that got downloaded because the usernames contain one of the key terms (i.e. “draghi”,¹ “ecb”, or—in the German sample—“ezb”) are in fact related to the ECB. Here, also users are excluded that are clearly connected to the English Cricket Board. This results in a list of around 300 users to disregard. Since it is common Twitter practice to mention users in tweets (preceded by a “@”), the tweets that contain the identified unrelated usernames in their text are also removed. This results in remaining ~3.5 million tweets in English and ~103,000 tweets in German.

To double-check for the language of tweets, the Python library “langdetect”² is used because despite the language filter of the Twitter Advanced Search, numerous tweets in other languages were returned. For the sample of tweets classified as English by Twitter, only those that “langdetect” also identifies as English are kept. For the sample of tweets classified as German by Twitter, detected languages may either be German or English due to the common usage of English terms even when the tweet is primarily in German language. This results in dropping around 5% of both, tweets in English and in German.

To understand different types of Twitter users and their behaviour, all tweets by users who have tweeted less than 100 times in their entire Twitter history are dropped. This leads to a loss of ~20,000 English tweets written by ~15,000 user accounts (~5% of all accounts in sample) and less than 1,000 German tweets by ~450 accounts (~3% of all accounts in sample), which has no impact on the time series properties of the variables that are subsequently studied.

B.1.2 Summary Statistics

Table B.1. Number of ECB-Related Tweets and Retweets

<i>Year</i>	<i>English sample</i>		<i>German sample</i>	
	<i>Tweets</i>	<i>Retweets</i>	<i>Tweets</i>	<i>Retweets</i>
2012	698,363	142,640	20,950	2,834
2013	428,683	123,059	10,946	1,840
2014	558,145	223,576	14,508	4,103
2015	650,872	497,894	17,256	7,275
2016	392,506	271,602	15,824	7,354
2017	277,777	208,673	10,333	4,885
2018	212,747	235,738	7,130	4,013
<i>Total</i>	3,219,093	1,703,182	96,947	32,304

Notes: The table shows the number of ECB-related tweets and retweets, by year. Tweets in English are reported in the left panel, tweets in German in the right panel.

¹Note that this is an Italian surname and thus not unlikely to occur in a username. In addition, it means “dragons” in Italian.

²Danilak (2015), available at <https://github.com/Mimino666/langdetect>. The langdetect library is a direct port of Google’s language-detection library, which generates language profiles from Wikipedia abstracts and claims to have 99% precision in language detection.

Table B.2. Descriptive Statistics of Tweet Content

<i>English sample</i>							
	<i>Mean</i>	<i>Std</i>	<i>Min</i>	<i>0.250</i>	<i>0.500</i>	<i>0.750</i>	<i>Max</i>
Subjectivity	0.247	0.279	0.000	0.000	0.150	0.455	1.000
Favourableness	0.043	0.197	-1.000	0.000	0.000	0.100	1.000
Absolute favourableness	0.107	0.171	0.000	0.000	0.000	0.1625	1.000

<i>German sample</i>							
	<i>Mean</i>	<i>Std</i>	<i>Min</i>	<i>0.250</i>	<i>0.500</i>	<i>0.750</i>	<i>Max</i>
Subjectivity	0.033	0.166	0.000	0.000	0.000	0.000	1.000
Favourableness	0.027	0.463	-1.000	0.000	0.000	0.000	1.000
Absolute favourableness	0.252	0.390	0.000	0.000	0.000	0.700	1.000

Notes: The table shows the descriptive statistics of the subjectivity, favourableness and absolute favourableness of English (top) and German (bottom) tweets. Std: standard deviation; Min: minimum; Max: maximum; 0.250: 25th percentile; 0.500: median; 0.750: 75th percentile.

Table B.3. Examples of Words in English Sentiment Lexicon

<i>selection of high in favourableness</i>				<i>selection of low in subjectivity</i>			
word	fav.	subj.	sense	word	fav.	subj.	sense
astonishing	1	1	so surprising as to stun or overwhelm	drag	-0.2	0	move slowly and as if with great effort
best	1	0.3	(superlative) having the most positive qualities	stretched	-0.1	0	extended spread over a wide area or distance
breathtaking	1	1	tending to cause suspension of regular breathing	unexplained	-0.1	0	having the reason or cause not made clear
consummate	1	1	having or revealing supreme mastery or skill	vacuum	-0.05	0	a region that is devoid of matter
delicious	1	1	extremely pleasing to the sense of taste	20th	0	0	coming next after the nineteenth in position
exceptional	1	1	surpassing what is common or usual or expected	academic	0	0	associated with academia or an academy
exceptional	1	1	far beyond what is usual in magnitude or degree	actual	0	0	being or existing at the present moment
impressed	1	1	deeply or markedly affected or influenced	aforementioned	0	0	being the one previously mentioned or spoken of
marvelous	1	1	too improbable to admit of belief	alternate	0	0	serving or used in place of another
marvelous	1	1	being or having the character of a miracle	atmospheric	0	0	relating to or located in the atmosphere
masterful	1	1	having or revealing supreme mastery or skill	back	0	0	relating to or located at the back
overwhelming	1	1	so strong as to be irresistible	basic	0	0	serving as a base or starting point
priceless	1	1	having incalculable monetary, intellectual or spiritual worth	basic	0	0	pertaining to or constituting a base or basis
bewitching	0.9	1	capturing interest as if by a spell	chronological	0	0	relating to or arranged according to temporal order
consummate	0.9	1	having or revealing supreme mastery or skill	comic	0	0	of or relating to or characteristic of comedy
favoured	0.8	0.9	preferred above all others and treated with partiality	consistent	0	0	the same throughout in structure or composition
fly	0.8	0.9	(British informal) not to be deceived or hoodwinked	contemporary	0	0	occurring in the same period of time
joy	0.8	0.2	something that provides a source of happiness	daily	0	0	of or belonging to or occurring every day

<i>selection of high in subjectivity</i>				<i>selection of low in favourableness</i>			
word	fav.	subj.	sense	word	fav.	subj.	sense
consummate	0.9	1	having or revealing supreme mastery or skill	awful	-1	1	causing fear or dread or terror
bewitching	0.7	1	capturing interest as if by a spell	deadly	-1	1	involving loss of divine grace or spiritual death
controversial	0.7	1	marked by or capable of arousing controversy	devastating	-1	1	wreaking or capable of wreaking complete destruction
astounding	0.6	1	bewildering or striking dumb with wonder	dreadful	-1	1	causing fear or dread or terror
bewitching	0.6	1	capturing interest as if by a spell	evil	-1	1	having or exerting a malignant influence
loving	0.6	1	feeling or showing love and affection	grim	-1	1	harshly uninviting or formidable in manner or appearance
mouth-watering	0.6	1	pleasing to the sense of taste	grotesque	-1	1	distorted and unnatural in shape or size
rose	0.6	1	of something having a dusty purplish pink color	horrific	-1	1	causing fear or dread or terror
adorable	0.5	1	lovely especially in a childlike or naive way	hysterical	-1	1	characterized by or arising from psychoneurotic behavior
authentic	0.5	1	conforming to fact and therefore worthy of belief	impossible	-1	1	used of persons or their behavior
avid	0.5	1	marked by active interest and enthusiasm	insane	-1	1	afflicted with or characterized of mental derangement
capable	0.5	1	have the skills and qualification to do things well	menacing	-1	1	threatening or foreshadow evil or tragic development
captivating	0.5	1	capturing interest as if by a spell	nasty	-1	1	exasperatingly difficult to handle or circumvent
certain	0.5	1	having or feeling no doubt or uncertainty	outrageous	-1	1	greatly exceeding bounds of reason or moderation
challenging	0.5	1	requiring full use of your abilities or resources	terrible	-1	1	causing fear or dread or terror
charismatic	0.5	1	possessing an extraordinary ability to attract	violent	-1	1	effected by force or injury rather than natural causes
competent	0.5	1	properly sufficiently qualified or capable or efficient	malevolent	-0.9	1	wishing or appearing to wish evil to others
confident	0.5	1	having or marked by confidence or assurance	repellent	-0.9	1	incapable of absorbing or mixing with
inconvenient	-0.6	1	not suited to your comfort, purpose or needs	stupid	-0.9	1	lacking or marked by lack of intellectual acuity

Notes: This table lists selected words and their favourableness and subjectivity scores. Multiple entries of the same word are generally due to multiple meanings and in these cases average score is taken by default. Source: Princeton University's WordNet, <https://wordnet.princeton.edu/>

Table B.4. Characteristics of Account Types, Robustness

	Number of accounts	Average date of account creation	Average percentile followers	Average percentile ECB centricity	Average subjectivity	Average standard deviation of subjectivity	Standard deviation of average subjectivity	Average favourable- ness	Average standard deviation of favourable- ness	Standard deviation of average favourable- ness	Average absolute favourable- ness	Average standard deviation of absolute favourable- ness	Standard deviation of average absolute favourable- ness	Average weekend activity
Panel A: English sample														
Non-expert (benchmark)	61,278	01/08/2011	68	12	0.2750	0.2163	0.2722	0.0515	0.1498	0.2161	0.1350	0.1285	0.1837	0.1837
Non-expert (non-centric)	254,178	29/12/2011	50	48	0.2842	0.2334	0.2503	0.0549	0.1652	0.1987	0.1362	0.1399	0.1691	0.1644
Non-expert (few followers)	4,610	12/06/2013	12	13	0.2271	0.1857	0.2542	0.0456	0.1193	0.1915	0.1091	0.1080	0.1676	0.1891
Expert (benchmark)	1,158	23/12/2012	68	84	0.2472	0.2578	0.0926	0.0428	0.1709	0.0567	0.0997	0.1483	0.0511	0.0691
Expert (0.33)	2,532	21/10/2012	66	82	0.2453	0.2554	0.0881	0.0426	0.1719	0.0502	0.1001	0.1488	0.0475	0.0714
Expert (0.75)	332	26/07/2013	67	87	0.2523	0.2653	0.1020	0.0481	0.1746	0.0731	0.1044	0.1519	0.0644	0.0563
Expert (centric)	968	06/10/2012	69	89	0.2487	0.2598	0.0859	0.0423	0.1732	0.0530	0.1005	0.1497	0.0490	0.0664
Panel B: German sample														
Non-expert (benchmark)	3,548	23/08/2011	65	12	0.0211	0.0335	0.1250	0.0274	0.2054	0.3354	0.1576	0.1802	0.3062	0.2026
Non-expert (non-centric)	14,762	26/01/2012	50	49	0.0313	0.0666	0.1333	0.0462	0.3435	0.3615	0.2311	0.2886	0.3166	0.1707
Non-expert (few followers)	310	24/08/2012	14	14	0.0056	0.0192	0.0612	-0.0015	0.1135	0.2599	0.0903	0.0985	0.2500	0.2560
Expert (benchmark)	18	22/01/2013	67	85	0.0603	0.1635	0.1129	0.0264	0.4242	0.2099	0.2612	0.3635	0.1394	0.0567
Expert (0.33)	69	04/03/2012	67	83	0.0419	0.1629	0.0606	0.0260	0.4462	0.1180	0.2518	0.3750	0.0982	0.0903
Expert (0.75)	4	26/10/2013	54	95	0.1448	0.1526	0.2369	-0.0948	0.4224	0.2703	0.2873	0.3641	0.1502	0.0126
Expert (centric)	15	17/10/2012	70	91	0.0692	0.1662	0.1222	-0.0147	0.4208	0.1376	0.2347	0.3613	0.0951	0.0526

Notes: The table shows summary statistics for the various account types, in the English sample (Panel A) and the German sample (Panel B). Statistics related to tweet sentiment exclude any tweets issued on press conference days.

Table B.5. Twitter Traffic, Leads and Lags of Press Conference and “Whatever It Takes”

	Log number of tweets					
	English			German		
	All	Non-experts	Experts	All	Non-experts	Experts
Press Conference, t-5	0.120* (0.071)	0.024 (0.089)	0.289*** (0.091)	--	--	--
Press Conference, t-4	0.244*** (0.088)	0.160 (0.103)	0.472*** (0.104)	--	--	--
Press Conference, t-3	0.227*** (0.073)	0.023 (0.097)	0.270*** (0.077)	--	--	--
Press Conference, t-2	0.212*** (0.077)	0.057 (0.103)	0.211*** (0.079)	--	--	--
Press Conference, t-1	0.569*** (0.082)	0.243** (0.096)	0.652*** (0.086)	0.434*** (0.112)	0.148 (0.152)	0.558*** (0.141)
Press Conference, t	2.467*** (0.083)	2.056*** (0.121)	2.816*** (0.086)	2.422*** (0.126)	1.200*** (0.181)	2.605*** (0.148)
Press Conference, t+1	0.976*** (0.090)	1.013*** (0.117)	0.903*** (0.093)	1.239*** (0.109)	0.596*** (0.159)	0.948*** (0.146)
Press Conference, t+2	0.399*** (0.086)	0.325*** (0.093)	0.476*** (0.108)	0.416*** (0.150)	0.092 (0.190)	0.185 (0.163)
Press Conference, t+3	0.234** (0.094)	0.216** (0.108)	0.270** (0.118)	--	--	--
Whatever it takes, t	1.975*** (0.080)	1.929*** (0.103)	1.831*** (0.090)	3.198*** (0.133)	1.530*** (0.180)	2.453*** (0.159)
Whatever it takes, t+1	2.802*** (0.071)	2.430*** (0.087)	2.839*** (0.083)	4.064*** (0.112)	2.529*** (0.142)	2.665*** (0.141)
Whatever it takes, t+2	1.809*** (0.080)	1.466*** (0.100)	1.473*** (0.094)	3.406*** (0.097)	2.274*** (0.121)	0.190 (0.139)
Whatever it takes, t+3	1.292*** (0.095)	1.333*** (0.109)	1.157*** (0.104)	1.968*** (0.097)	0.953*** (0.127)	1.110*** (0.135)
Whatever it takes, t+4	1.825*** (0.084)	1.712*** (0.108)	1.614*** (0.089)	3.852*** (0.094)	3.581*** (0.123)	2.286*** (0.120)
Whatever it takes, t+5	1.940*** (0.094)	1.783*** (0.118)	1.698*** (0.101)	3.420*** (0.108)	2.863*** (0.133)	2.350*** (0.130)
Whatever it takes, t+6	1.439*** (0.097)	1.356*** (0.105)	1.379*** (0.108)	2.382*** (0.141)	1.835*** (0.185)	1.353*** (0.177)
Whatever it takes, t+7	1.407*** (0.087)	1.614*** (0.115)	1.141*** (0.094)	2.753*** (0.131)	2.538*** (0.183)	1.130*** (0.154)
Whatever it takes, t+8	1.669*** (0.109)	1.531*** (0.131)	1.636*** (0.120)	2.823*** (0.145)	2.088*** (0.201)	1.290*** (0.188)
Whatever it takes, t+9	1.487*** (0.098)	1.420*** (0.101)	1.700*** (0.119)	3.368*** (0.170)	2.114*** (0.220)	1.025*** (0.185)
Whatever it takes, t+10	1.315*** (0.103)	1.004*** (0.112)	1.290*** (0.125)	2.662*** (0.113)	1.910*** (0.137)	--
Whatever it takes, t+11	1.643*** (0.070)	1.125*** (0.074)	1.627*** (0.084)	2.951*** (0.110)	2.547*** (0.132)	-0.358** (0.144)
Whatever it takes, t+12	1.613*** (0.073)	1.213*** (0.080)	1.343*** (0.088)	2.717*** (0.119)	0.379** (0.149)	--
Whatever it takes, t+13	0.953*** (0.070)	0.243*** (0.072)	0.923*** (0.084)	1.522*** (0.109)	-0.034 (0.136)	--
Whatever it takes, t+14	1.362*** (0.094)	0.688*** (0.109)	1.554*** (0.105)	2.011*** (0.149)	1.088*** (0.193)	--
Whatever it takes, t+15	0.698*** (0.072)	0.461*** (0.073)	0.507*** (0.087)	0.747*** (0.114)	-0.726*** (0.146)	--
Whatever it takes, t+16	0.759*** (0.071)	0.746*** (0.075)	0.370*** (0.093)	--	--	--
Whatever it takes, t+17	1.389*** (0.074)	2.190*** (0.075)	0.368*** (0.097)	--	--	--
Whatever it takes, t+18	0.826*** (0.070)	1.132*** (0.074)	0.566*** (0.084)	--	--	--
Whatever it takes, t+19	1.052*** (0.073)	1.258*** (0.080)	1.119*** (0.088)	--	--	--
Whatever it takes, t+20	0.788*** (0.077)	1.243*** (0.083)	0.462*** (0.092)	--	--	--

Notes: The table shows the coefficient estimates for leads and lags of the ECB press conference and “Whatever it takes” omitted from Table 3. Numbers in brackets are standard errors. ***/**/* denote statistical significance at the 1%/5%/10% level.

B.2 Robustness of Results

Appendix Tables B.6-B.13 provide the estimated coefficients for the different ways of classifying non-experts and experts, for all dependent variables. Overall, results are remarkably robust. It is important to note that some of the groups are rather small—this is in particular the case for the most restricted definition of non-experts in the sample of German tweets. It comprises 310 accounts, from which only few tweets are posted, such that there are only 254 observations at the daily aggregate level. This needs to be kept in mind when studying the results.

With regard to Twitter volume, the main findings (i) non-experts are less responsive than experts, in particular to some, more specialised, types of ECB communication; ii) the smaller responsiveness is also reflected in a lower R²; iii) the exception to this is the “Whatever it takes” statement, which led to a similar response by non-experts and experts) all go through, independent of the exact way of defining experts and non-experts.

Coming to subjectivity, most results are also confirmed. However, mean subjectivity is not responsive to “Whatever it takes” on the day of the statement, for both robustness tests for non-experts. The other results go through: the standard deviation of subjectivity increases for the experts following the press conference, and in the German sample, the standard deviation of subjectivity increases in response to “Whatever it takes”, both on the same day and over the duration of the Twitter discussion, both for experts and non-experts.

For favourableness, the main results were a decrease in its standard deviation for the tweets in English, for experts and non-experts alike, and an increase for the tweets in German, in particular for “Whatever it takes”. The latter finding is robustly repeated across the various definitions. Also the former is broadly robust, once more with the exception of non-experts that have few followers, where the standard deviation of favourableness is not responsive to the press conference, whereas it is decreasing for the other non-expert groups.

Also for the last set of results, studying absolute favourableness, results are broadly robust, with the partial exception of English-speaking non-experts with few followers, where the standard deviation is not declining in response to the press conference.

To summarise, the robustness tests broadly confirm the earlier picture, but suggest that the group of non-experts with few followers in the English-speaking group behaves somewhat differently from other non-experts. Note that this group does not look any different per se in terms of the underlying characteristics (see Table B.4).

Table B.6. Twitter Traffic, Robustness Tests

Log number of tweets																
English sample								German sample								
	Non-experts (bm)	Non-experts (non-centr.)	Non-experts (few follow.)	Experts (bm)	Experts (0.33)	Experts (0.75)	Experts (centric)		Non-experts (bm)	Non-experts (excl centr.)	Non-experts (few follow.)	Experts (bm)	Experts (0.33)	Experts (0.75)	Experts (ECB centric)	
Panel A: Contemporaneous response																
Press Conference	2.056*** (0.121)	2.291*** (0.090)	1.829*** (0.130)	2.816*** (0.086)	2.809*** (0.084)	2.816*** (0.088)	2.803*** (0.086)		1.200*** (0.181)	2.311*** (0.128)	0.030 (0.235)	2.605*** (0.148)	2.948*** (0.158)	2.148*** (0.153)	2.452*** (0.152)	
Whatever it takes	1.929*** (0.103)	2.052*** (0.081)	1.967*** (0.130)	1.831*** (0.090)	1.907*** (0.089)	1.872*** (0.096)	1.818*** (0.089)		1.530*** (0.180)	3.225*** (0.134)	-0.189 (0.265)	2.453*** (0.159)	2.562*** (0.174)	1.992*** (0.183)	2.272*** (0.157)	
Economic Bulletin	0.180 (0.109)	0.184** (0.091)	-0.029 (0.126)	0.275*** (0.095)	0.257*** (0.095)	0.251*** (0.096)	0.279*** (0.094)		-0.264 (0.186)	-0.217 (0.140)	-0.514** (0.205)	-0.171 (0.162)	-0.276 (0.187)	-0.081 (0.157)	-0.222 (0.156)	
Accounts	0.312*** (0.098)	0.391*** (0.092)	0.124 (0.155)	0.853*** (0.095)	0.811*** (0.095)	0.833*** (0.100)	0.849*** (0.094)		0.067 (0.214)	0.073 (0.161)	0.027 (0.471)	-0.146 (0.186)	-0.159 (0.232)	-0.065 (0.213)	-0.224 (0.179)	
Speeches by others	0.084 (0.057)	0.202*** (0.045)	0.013 (0.070)	0.394*** (0.053)	0.364*** (0.051)	0.351*** (0.058)	0.392*** (0.053)		0.031 (0.097)	0.146* (0.076)	-0.018 (0.182)	0.044 (0.093)	0.158 (0.103)	0.080 (0.098)	0.002 (0.094)	
Speeches by president	0.360*** (0.074)	0.416*** (0.055)	0.375*** (0.099)	0.509*** (0.057)	0.518*** (0.058)	0.535*** (0.058)	0.507*** (0.056)		0.488*** (0.132)	0.788*** (0.096)	0.316 (0.195)	1.119*** (0.104)	1.135*** (0.121)	0.906*** (0.106)	1.107*** (0.103)	
Tweet	0.150*** (0.054)	0.173*** (0.044)	0.047 (0.062)	0.247*** (0.053)	0.241*** (0.051)	0.186*** (0.060)	0.244*** (0.053)		0.076 (0.087)	0.097 (0.070)	-0.175 (0.150)	0.075 (0.087)	0.106 (0.096)	0.075 (0.090)	0.052 (0.087)	
Panel B: Overall response																
Press Conference	4.116	5.176	3.572	6.358	6.373	6.198	6.302		2.036	4.334	0.976	4.295	4.879	2.688	4.061	
Std. error	0.335	0.274	0.407	0.315	0.302	0.346	0.313		0.350	0.257	0.851	0.308	0.367	0.311	0.313	
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.253	0.000	0.000	0.000	0.000	
Whatever it takes	27.876	30.945	18.327	26.597	28.421	25.365	26.529		27.471	43.700	1.618	15.494	23.044	6.202	14.896	
Std. error	1.052	1.046	1.222	1.276	1.244	1.336	1.266		1.518	1.223	1.921	1.008	1.671	0.892	0.979	
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.401	0.000	0.000	0.000	0.000	
Observations	2,532	2,532	1,972	2,532	2,532	2,527	2,532		1,461	2,522	254	1,177	1,516	679	1,139	
R-squared	0.324	0.535	0.270	0.628	0.643	0.619	0.632		0.177	0.333	0.211	0.442	0.356	0.498	0.441	

Notes: The table shows coefficient estimates for the effect of ECB communication events on log number of tweets, based on Equation (2.6), for different definitions of non-experts and experts. The models control for day of week, month of year and holiday effects and allow for a linear and quadratic time trend. They contain 5 (1) leads and 3 (2) lags for the press conference and 20 (15) lags for “Whatever it takes” in the English (German) sample. Panel B reports the accumulated response coefficients over all leads and lags, with coefficients that are statistically significant at least at the 10% level printed in bold. Numbers in brackets are standard errors. ***/**/* denote statistical significance at the 1%/5%/10% level.

Table B.7. Concentration Index, Robustness Tests

Concentration index														
	English sample							German sample						
	Non-experts (bm)	Non-experts (non-centr.)	Non-experts (few follow.)	Experts (bm)	Experts (0.33)	Experts (0.75)	Experts (centric)	Non-experts (bm)	Non-experts (excl centr.)	Non-experts (few follow.)	Experts (bm)	Experts (0.33)	Experts (0.75)	Experts (ECB centric)
Panel A: Contemporaneous response														
Press Conference	-0.043*** (0.003)	-0.005*** (0.001)	-0.385*** (0.034)	-0.024*** (0.003)	-0.020*** (0.002)	-0.051*** (0.007)	-0.025*** (0.003)	-0.403*** (0.047)	-0.139*** (0.016)	-0.077 (0.104)	-0.538*** (0.041)	-0.476*** (0.040)	-0.459*** (0.036)	-0.516*** (0.042)
Whatever it takes	-0.017*** (0.004)	-0.003*** (0.001)	-0.218*** (0.045)	-0.013*** (0.004)	-0.011*** (0.003)	-0.028*** (0.008)	-0.013*** (0.004)	-0.430*** (0.059)	-0.102*** (0.019)	-0.237** (0.118)	-0.419*** (0.053)	-0.329*** (0.053)	-0.416*** (0.051)	-0.403*** (0.053)
Economic Bulletin	-0.006* (0.004)	-0.001 (0.001)	0.012 (0.048)	-0.008*** (0.003)	-0.006*** (0.002)	-0.020*** (0.006)	-0.008*** (0.003)	0.067 (0.073)	0.004 (0.025)	0.197** (0.087)	-0.014 (0.060)	0.042 (0.061)	-0.055 (0.051)	0.006 (0.059)
Accounts	-0.021*** (0.004)	-0.001* (0.001)	-0.080 (0.065)	-0.017*** (0.004)	-0.013*** (0.003)	-0.031*** (0.007)	-0.018*** (0.004)	-0.048 (0.075)	-0.032 (0.025)	0.008 (0.151)	0.005 (0.064)	0.030 (0.074)	0.019 (0.060)	0.023 (0.063)
Speeches by others	-0.006** (0.003)	-0.002** (0.001)	-0.000 (0.026)	-0.015*** (0.003)	-0.010*** (0.002)	-0.032*** (0.006)	-0.016*** (0.003)	-0.009 (0.034)	-0.038*** (0.014)	-0.016 (0.073)	-0.031 (0.034)	-0.069** (0.033)	-0.066*** (0.025)	-0.022 (0.034)
Speeches by president	-0.013*** (0.003)	-0.001*** (0.000)	-0.119*** (0.034)	-0.001 (0.001)	-0.002* (0.001)	-0.004 (0.003)	-0.001 (0.002)	-0.183*** (0.040)	-0.058*** (0.010)	-0.143 (0.088)	-0.305*** (0.030)	-0.253*** (0.030)	-0.224*** (0.030)	-0.302*** (0.031)
Tweet	-0.008*** (0.003)	-0.002** (0.001)	0.005 (0.024)	-0.012*** (0.003)	-0.008*** (0.002)	-0.024*** (0.007)	-0.013*** (0.003)	-0.040 (0.031)	-0.030** (0.014)	0.019 (0.065)	-0.051 (0.032)	-0.052* (0.031)	-0.062*** (0.020)	-0.056* (0.032)
Panel B: Overall response														
Press Conference	-0.138	-0.025	-1.045	-0.226	-0.147	-0.495	-0.241	-0.776	-0.367	-0.530	-1.061	-0.952	-0.600	-0.997
Std. error	0.016	0.003	0.161	0.023	0.014	0.047	0.024	0.112	0.034	0.280	0.109	0.105	0.080	0.111
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.060	0.000	0.000	0.000	0.000
Whatever it takes	-0.720	-0.095	-3.919	-0.875	-0.287	-2.020	-0.920	-6.962	-3.107	-1.773	-3.911	-5.524	-1.428	-3.547
Std. error	0.081	0.013	0.472	0.099	0.072	0.202	0.102	0.516	0.212	0.744	0.354	0.535	0.210	0.329
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000	0.000
Observations	2,532	2,532	1,972	2,532	2,532	2,527	2,532	1,461	2,522	254	1,177	1,516	679	1,139
R-squared	0.198	0.188	0.174	0.394	0.375	0.420	0.408	0.167	0.159	0.215	0.277	0.190	0.408	0.278

Notes: The table shows coefficient estimates for the effect of ECB communication events on the concentration index, based on Equation (2.6), for different definitions of non-experts and experts. The models control for day of week, month of year and holiday effects and allow for a linear and quadratic time trend. They contain 5 (1) leads and 3 (2) lags for the press conference and 20 (15) lags for “Whatever it takes” in the English (German) sample. Panel B reports the accumulated response coefficients over all leads and lags, with coefficients that are statistically significant at least at the 10% level printed in bold. Numbers in brackets are standard errors. ***/**/* denote statistical significance at the 1%/5%/10% level.

Table B.8. Average Subjectivity, Robustness Tests

Average subjectivity														
	English sample							German sample						
	Non-experts (non-centr.) (bm)	Non-experts (few centr.) (bm)	Non-experts (few follow.) (bm)	Experts (bm)	Experts (0.33) (0.33)	Experts (0.75) (0.75)	Experts (centric) (centric)	Non-experts (non-centr.) (bm)	Non-experts (excl centr.) (bm)	Non-experts (few follow.) (bm)	Experts (bm)	Experts (0.33) (0.33)	Experts (0.75) (0.75)	Experts (ECB centric) (centric)
Panel A: Contemporaneous response														
Press Conference	-0.031*** (0.007)	-0.015*** (0.005)	-0.029*** (0.010)	0.012* (0.006)	0.015*** (0.006)	0.023*** (0.007)	0.019*** (0.005)	-0.031* (0.017)	-0.010 (0.006)	-0.022 (0.038)	-0.023 (0.022)	-0.006 (0.016)	-0.073* (0.040)	-0.028 (0.023)
Whatever it takes	-0.021** (0.008)	-0.009 (0.006)	-0.014 (0.012)	0.016** (0.008)	0.013* (0.007)	0.011 (0.009)	0.016** (0.007)	0.040* (0.023)	-0.011 (0.008)	0.019 (0.028)	-0.001 (0.029)	0.009 (0.019)	0.032 (0.045)	0.006 (0.031)
Economic Bulletin	-0.008 (0.009)	-0.012* (0.006)	-0.013 (0.012)	-0.002 (0.007)	-0.005 (0.006)	0.004 (0.007)	-0.004 (0.006)	-0.020 (0.033)	-0.007 (0.008)	-0.047 (0.041)	-0.001 (0.028)	0.027 (0.023)	-0.034 (0.060)	-0.001 (0.031)
Accounts	-0.025** (0.011)	-0.026*** (0.006)	-0.016 (0.018)	-0.015* (0.007)	-0.008 (0.007)	-0.013 (0.008)	-0.015** (0.007)	-0.054*** (0.017)	-0.001 (0.011)	-0.018 (0.027)	-0.055** (0.022)	-0.043*** (0.015)	-0.096** (0.038)	-0.059*** (0.023)
Speeches by others	-0.002 (0.005)	-0.001 (0.004)	-0.006 (0.007)	0.008 (0.005)	0.009* (0.005)	0.010* (0.006)	0.007 (0.004)	-0.004 (0.014)	-0.004 (0.004)	-0.055 (0.047)	-0.010 (0.015)	-0.015 (0.013)	-0.027 (0.031)	-0.013 (0.016)
Speeches by president	-0.018*** (0.006)	-0.008* (0.004)	-0.018** (0.009)	-0.005 (0.005)	-0.002 (0.004)	0.000 (0.005)	0.002 (0.004)	-0.002 (0.015)	0.002 (0.006)	-0.002 (0.024)	-0.011 (0.011)	-0.001 (0.011)	-0.014 (0.018)	-0.010 (0.011)
Tweet	-0.006 (0.004)	-0.005 (0.003)	-0.008 (0.006)	-0.002 (0.005)	-0.003 (0.004)	-0.002 (0.005)	-0.003 (0.004)	-0.010 (0.013)	-0.004 (0.004)	-0.061 (0.050)	0.011 (0.015)	-0.000 (0.013)	-0.013 (0.032)	0.008 (0.016)
Panel B: Overall response														
Press Conference	-0.115	-0.079	-0.079	0.005	0.036	0.023	0.038	-0.008	-0.024	-0.002	0.017	0.002	0.014	0.004
Std. error	0.028	0.022	0.041	0.032	0.029	0.037	0.032	0.041	0.014	0.068	0.053	0.042	0.085	0.054
p-value	0.000	0.000	0.055	0.871	0.220	0.528	0.231	0.845	0.090	0.981	0.743	0.959	0.866	0.945
Whatever it takes	0.096	0.090	0.276	0.544	0.442	0.757	0.731	0.202	0.043	1.071	-0.146	-0.035	-0.190	-0.096
Std. error	0.092	0.076	0.137	0.104	0.095	0.131	0.105	0.154	0.065	0.208	0.169	0.212	0.173	0.162
p-value	0.297	0.238	0.044	0.000	0.000	0.000	0.000	0.190	0.501	0.000	0.388	0.870	0.273	0.556
Observations	2,532	2,532	2,532	2,532	2,532	2,527	2,531	1,461	2,522	254	1,177	1,516	679	1,139
R-squared	0.145	0.183	0.105	0.110	0.075	0.147	0.183	0.031	0.036	0.410	0.048	0.047	0.047	0.047

Notes: The table shows coefficient estimates for the effect of ECB communication events on average subjectivity, based on Equation (2.6), for different definitions of non-experts and experts. The models control for day of week, month of year and holiday effects and allow for a linear and quadratic time trend. They contain 5 (1) leads and 3 (2) lags for the press conference and 20 (15) lags for “Whatever it takes” in the English (German) sample. Panel B reports the accumulated response coefficients over all leads and lags, with coefficients that are statistically significant at least at the 10% level printed in bold. Numbers in brackets are standard errors. ***/**/* denote statistical significance at the 1%/5%/10% level.

Table B.9. Standard Deviation of Subjectivity, Robustness Tests

Standard deviation of subjectivity														
	English sample							German sample						
	Non-experts (bm)	Non-experts (non-centr.)	Non-experts (few follow.)	Experts (bm)	Experts (0.33)	Experts (0.75)	Experts (centric)	Non-experts (bm)	Non-experts (excl centr.)	Non-experts (few follow.)	Experts (bm)	Experts (0.33)	Experts (0.75)	Experts (ECB centric)
Panel A: Contemporaneous response														
Press Conference	-0.022*** (0.004)	-0.013*** (0.003)	-0.017*** (0.006)	0.002 (0.004)	-0.001 (0.003)	0.008* (0.005)	0.008** (0.004)	0.008 (0.017)	0.033** (0.013)	0.025 (0.023)	0.069*** (0.019)	0.089*** (0.016)	0.055*** (0.018)	0.061*** (0.019)
Whatever it takes	-0.013*** (0.004)	-0.005 (0.003)	-0.006 (0.007)	0.006 (0.005)	0.005 (0.005)	0.004 (0.006)	0.005 (0.005)	0.204*** (0.018)	0.000 (0.016)	-0.009 (0.017)	0.157*** (0.016)	0.087*** (0.018)	0.258*** (0.019)	0.178*** (0.017)
Economic Bulletin	-0.010** (0.004)	-0.010*** (0.003)	-0.007 (0.007)	-0.006 (0.004)	-0.010** (0.004)	-0.005 (0.005)	-0.007 (0.004)	-0.036*** (0.012)	-0.010 (0.017)	-0.016 (0.024)	0.014 (0.025)	0.045* (0.026)	-0.020 (0.012)	0.012 (0.025)
Accounts	-0.011 (0.007)	-0.011** (0.004)	-0.005 (0.010)	-0.006 (0.006)	-0.008 (0.006)	-0.002 (0.007)	-0.002 (0.005)	-0.036*** (0.012)	0.014 (0.026)	-0.001 (0.009)	-0.027** (0.012)	-0.029*** (0.011)	-0.015 (0.012)	-0.032*** (0.012)
Speeches by others	-0.002 (0.003)	-0.001 (0.002)	-0.002 (0.004)	0.004 (0.003)	0.002 (0.003)	0.009** (0.004)	0.007** (0.003)	-0.011 (0.008)	-0.004 (0.009)	-0.002 (0.006)	-0.007 (0.012)	-0.003 (0.009)	0.001 (0.010)	-0.013 (0.012)
Speeches by president	-0.007** (0.003)	-0.004* (0.002)	-0.009* (0.004)	-0.004 (0.003)	-0.003 (0.003)	-0.002 (0.003)	-0.002 (0.003)	0.025* (0.013)	0.031*** (0.011)	0.019 (0.012)	0.017 (0.011)	0.029** (0.012)	0.021* (0.011)	0.019* (0.011)
Tweet	-0.004 (0.002)	-0.003 (0.002)	-0.002 (0.004)	-0.001 (0.003)	-0.004 (0.002)	0.000 (0.004)	0.002 (0.003)	-0.006 (0.008)	-0.008 (0.008)	0.007 (0.006)	0.002 (0.012)	0.004 (0.009)	0.013 (0.009)	-0.004 (0.012)
Panel B: Overall response														
Press Conference	-0.046	-0.023	0.016	0.063	0.048	0.134	0.081	0.057	0.069	0.058	0.169	0.185	0.180	0.153
Std. error	0.015	0.012	0.024	0.020	0.018	0.027	0.022	0.033	0.031	0.048	0.041	0.039	0.057	0.041
p-value	0.002	0.060	0.519	0.002	0.009	0.000	0.000	0.081	0.023	0.236	0.000	0.000	0.002	0.000
Whatever it takes	0.018	-0.078	0.263	0.076	-0.112	0.160	0.122	1.188	0.830	0.414	0.570	1.160	0.369	0.635
Std. error	0.049	0.039	0.080	0.070	0.063	0.097	0.076	0.090	0.124	0.096	0.108	0.151	0.080	0.106
p-value	0.715	0.043	0.001	0.276	0.075	0.098	0.111	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	2,532	2,532	2,532	2,532	2,532	2,527	2,531	1,461	2,522	254	1,177	1,516	679	1,139
R-squared	0.067	0.072	0.046	0.102	0.072	0.176	0.167	0.062	0.077	0.307	0.094	0.087	0.102	0.096

Notes: The table shows coefficient estimates for the effect of ECB communication events on the standard deviation of subjectivity, based on Equation (2.6), for different definitions of non-experts and experts. The models control for day of week, month of year and holiday effects and allow for a linear and quadratic time trend. They contain 5 (1) leads and 3 (2) lags for the press conference and 20 (15) lags for “Whatever it takes” in the English (German) sample. Panel B reports the accumulated response coefficients over all leads and lags, with coefficients that are statistically significant at least at the 10% level printed in bold. Numbers in brackets are standard errors. ***/**/* denote statistical significance at the 1%/5%/10% level.

Table B.10. Average Favourableness, Robustness Tests

	Average favourableness													
	English sample							German sample						
	Non-experts (non-centr.) (bm)	Non-experts (few follow.)	Experts (bm)	Experts (0.33)	Experts (0.75)	Experts (centric)		Non-experts (bm)	Non-experts (excl centr.)	Non-experts (few follow.)	Experts (bm)	Experts (0.33)	Experts (0.75)	Experts (ECB centric)
Panel A: Contemporaneous response														
Press Conference	-0.007 (0.006)	-0.002 (0.004)	-0.008 (0.008)	0.004 (0.004)	0.007* (0.004)	0.003 (0.004)	-0.001 (0.003)	-0.037 (0.048)	-0.030 (0.020)	-0.003 (0.093)	0.025 (0.049)	0.028 (0.042)	0.078 (0.087)	0.021 (0.052)
Whatever it takes	0.008 (0.007)	0.016*** (0.005)	0.006 (0.009)	0.017*** (0.005)	0.022*** (0.005)	0.013** (0.005)	0.017*** (0.005)	0.054 (0.051)	0.050** (0.025)	0.235 (0.177)	0.096 (0.059)	0.055 (0.054)	0.070 (0.091)	0.116* (0.063)
Economic Bulletin	0.008 (0.006)	0.006 (0.004)	0.002 (0.009)	0.005 (0.004)	0.007* (0.004)	0.004 (0.005)	0.000 (0.004)	0.001 (0.050)	-0.029 (0.030)	-0.029 (0.114)	0.043 (0.058)	0.022 (0.055)	0.046 (0.107)	0.052 (0.061)
Accounts	0.010 (0.007)	0.007 (0.005)	0.009 (0.011)	0.005 (0.005)	0.009** (0.005)	0.002 (0.006)	-0.002 (0.005)	0.077 (0.067)	-0.032 (0.037)	0.323 (0.235)	-0.032 (0.064)	0.022 (0.059)	0.111 (0.088)	-0.032 (0.068)
Speeches by others	-0.001 (0.004)	0.002 (0.003)	0.000 (0.005)	0.002 (0.003)	0.003 (0.003)	0.001 (0.003)	0.001 (0.003)	0.043 (0.032)	-0.017 (0.014)	-0.001 (0.087)	-0.040 (0.040)	0.006 (0.035)	-0.079 (0.074)	-0.051 (0.043)
Speeches by president	0.004 (0.004)	0.004 (0.003)	0.009* (0.006)	0.007** (0.003)	0.007** (0.003)	0.006* (0.003)	0.006** (0.003)	-0.090*** (0.030)	-0.004 (0.015)	0.000 (0.088)	0.036 (0.031)	0.017 (0.030)	0.046 (0.043)	0.043 (0.033)
Tweet	-0.002 (0.003)	-0.000 (0.003)	-0.002 (0.005)	0.000 (0.003)	0.001 (0.003)	-0.001 (0.003)	-0.002 (0.003)	0.046* (0.027)	-0.007 (0.012)	0.115 (0.089)	-0.009 (0.039)	0.026 (0.033)	-0.033 (0.074)	-0.016 (0.042)
Panel B: Overall response														
Press Conference	0.003	-0.013	0.022	0.025	0.025	0.012	0.008	-0.171	-0.049	-0.376	-0.052	-0.069	0.270	-0.093
Std. error	0.021	0.017	0.030	0.020	0.019	0.022	0.017	0.102	0.048	0.269	0.134	0.112	0.207	0.146
p-value	0.888	0.454	0.460	0.219	0.201	0.576	0.642	0.094	0.316	0.163	0.701	0.539	0.193	0.522
Whatever it takes	-0.197	-0.122	-0.268	0.021	0.133	0.122	0.046	-0.512	-0.951	1.267	-1.905	-1.939	0.030	-1.847
Std. error	0.066	0.053	0.101	0.063	0.058	0.075	0.059	0.399	0.223	1.006	0.466	0.572	0.481	0.466
p-value	0.003	0.021	0.008	0.736	0.022	0.101	0.436	0.199	0.000	0.209	0.000	0.001	0.951	0.000
Observations	2,532	2,532	2,532	2,532	2,532	2,527	2,531	1,461	2,522	254	1,177	1,516	679	1,139
R-squared	0.053	0.075	0.039	0.043	0.050	0.035	0.059	0.047	0.034	0.167	0.050	0.041	0.088	0.052

Notes: The table shows coefficient estimates for the effect of ECB communication events on average favourableness, based on Equation (2.6), for different definitions of non-experts and experts. The models control for day of week, month of year and holiday effects and allow for a linear and quadratic time trend. They contain 5 (1) leads and 3 (2) lags for the press conference and 20 (15) lags for “Whatever it takes” in the English (German) sample. Panel B reports the accumulated response coefficients over all leads and lags, with coefficients that are statistically significant at least at the 10% level printed in bold. Numbers in brackets are standard errors. ***/**/* denote statistical significance at the 1%/5%/10% level.

Table B.11. Standard Deviation of Favourableness, Robustness Tests

Standard deviation of favourableness														
	English sample							German sample						
	Non-experts (non-centr.) (bm)	Non-experts (few follow.)	Experts (bm)	Experts (0.33)	Experts (0.75)	Experts (centric)		Non-experts (excl centr.) (bm)	Non-experts (few follow.)	Experts (bm)	Experts (0.33)	Experts (0.75)	Experts (ECB centric)	
Panel A: Contemporaneous response														
Press Conference	-0.045*** (0.005)	-0.031*** (0.004)	-0.045*** (0.007)	-0.010** (0.004)	-0.010** (0.004)	-0.002 (0.005)	-0.001 (0.004)	0.201*** (0.037)	0.080*** (0.017)	0.028 (0.032)	0.272*** (0.030)	0.278*** (0.027)	0.287*** (0.036)	0.254*** (0.031)
Whatever it takes	-0.030*** (0.006)	-0.014*** (0.004)	-0.032*** (0.008)	0.016*** (0.005)	0.015*** (0.005)	0.021*** (0.006)	0.029*** (0.005)	0.309*** (0.033)	0.159*** (0.020)	-0.228* (0.133)	0.297*** (0.039)	0.203*** (0.036)	0.264*** (0.048)	0.253*** (0.039)
Economic Bulletin	-0.003 (0.006)	-0.008* (0.005)	-0.004 (0.008)	-0.005 (0.005)	-0.007 (0.005)	-0.007 (0.005)	-0.008 (0.005)	-0.022 (0.036)	0.004 (0.023)	-0.017 (0.034)	0.013 (0.044)	-0.011 (0.040)	0.060 (0.063)	0.010 (0.045)
Accounts	-0.017** (0.008)	-0.016*** (0.005)	-0.009 (0.011)	-0.015*** (0.005)	-0.013** (0.005)	-0.009 (0.006)	-0.012** (0.005)	0.043 (0.059)	-0.003 (0.028)	-0.044 (0.038)	-0.044 (0.035)	-0.006 (0.045)	-0.045 (0.038)	-0.041 (0.037)
Speeches by others	-0.005 (0.003)	-0.003 (0.003)	-0.011** (0.005)	0.003 (0.003)	-0.000 (0.003)	0.006 (0.004)	0.003 (0.003)	0.006 (0.019)	0.013 (0.013)	-0.002 (0.016)	0.024 (0.024)	0.048** (0.022)	0.039 (0.026)	0.023 (0.025)
Speeches by president	-0.014*** (0.004)	-0.010*** (0.003)	-0.016*** (0.006)	-0.010*** (0.003)	-0.007** (0.003)	-0.009** (0.003)	-0.007** (0.003)	0.055** (0.025)	0.039*** (0.014)	0.122** (0.057)	0.157*** (0.025)	0.142*** (0.024)	0.143*** (0.028)	0.163*** (0.025)
Tweet	-0.004 (0.003)	-0.002 (0.003)	-0.011*** (0.004)	-0.003 (0.003)	-0.004 (0.003)	-0.002 (0.004)	-0.003 (0.003)	0.013 (0.016)	0.024** (0.011)	-0.002 (0.014)	0.009 (0.022)	0.037* (0.020)	0.041* (0.023)	0.009 (0.023)
Panel B: Overall response														
Press Conference	-0.089	-0.101	-0.024	-0.005	-0.016	0.070	-0.006	0.304	0.171	0.104	0.575	0.484	0.419	0.542
Std. error	0.020	0.016	0.030	0.022	0.020	0.025	0.022	0.066	0.036	0.069	0.081	0.067	0.095	0.081
p-value	0.000	0.000	0.412	0.811	0.429	0.006	0.797	0.000	0.000	0.136	0.000	0.000	0.000	0.000
Whatever it takes	0.005	0.119	0.104	0.427	0.363	0.376	0.439	2.581	1.720	-0.115	1.676	2.323	0.809	1.753
Std. error	0.062	0.052	0.093	0.067	0.062	0.083	0.073	0.259	0.190	0.647	0.257	0.318	0.221	0.236
p-value	0.932	0.022	0.265	0.000	0.000	0.000	0.000	0.000	0.000	0.859	0.000	0.000	0.000	0.000
Observations	2,532	2,532	2,532	2,532	2,532	2,527	2,531	1,461	2,522	254	1,177	1,516	679	1,139
R-squared	0.096	0.140	0.060	0.105	0.068	0.147	0.136	0.097	0.088	0.231	0.181	0.142	0.245	0.190

Notes: The table shows coefficient estimates for the effect of ECB communication events on the standard deviation of favourableness, based on Equation (2.6), for different definitions of non-experts and experts. The models control for day of week, month of year and holiday effects and allow for a linear and quadratic time trend. They contain 5 (1) leads and 3 (2) lags for the press conference and 20 (15) lags for “Whatever it takes” in the English (German) sample. Panel B reports the accumulated response coefficients over all leads and lags, with coefficients that are statistically significant at least at the 10% level printed in bold. Numbers in brackets are standard errors. ***/**/* denote statistical significance at the 1%/5%/10% level.

Table B.12. Average Absolute Favourableness, Robustness Tests

Average absolute favourableness														
	English sample							German sample						
	Non-experts (bm)	Non-experts (non-centr.)	Non-experts (few follow.)	Experts (bm)	Experts (0.33)	Experts (0.75)	Experts (centric)	Non-experts (bm)	Non-experts (excl centr.)	Non-experts (few follow.)	Experts (bm)	Experts (0.33)	Experts (0.75)	Experts (ECB centric)
Panel A: Contemporaneous response														
Press Conference	-0.039*** (0.005)	-0.024*** (0.004)	-0.046*** (0.007)	-0.008** (0.003)	-0.007** (0.003)	-0.001 (0.004)	-0.001 (0.003)	0.101** (0.043)	0.035** (0.016)	0.101 (0.096)	0.013 (0.041)	0.020 (0.034)	-0.109 (0.069)	0.002 (0.043)
Whatever it takes	-0.021*** (0.006)	-0.008* (0.004)	-0.023*** (0.008)	0.011*** (0.004)	0.011** (0.004)	0.013*** (0.005)	0.019*** (0.004)	0.094** (0.046)	0.091*** (0.021)	-0.140 (0.200)	0.154*** (0.053)	0.032 (0.047)	-0.044 (0.079)	0.106* (0.055)
Economic Bulletin	-0.001 (0.006)	-0.005 (0.004)	-0.006 (0.008)	-0.002 (0.004)	-0.003 (0.004)	-0.001 (0.004)	-0.003 (0.003)	-0.036 (0.050)	0.019 (0.026)	-0.032 (0.137)	-0.017 (0.054)	-0.042 (0.048)	0.023 (0.084)	-0.005 (0.056)
Accounts	-0.016** (0.007)	-0.015*** (0.004)	-0.017* (0.009)	-0.012*** (0.004)	-0.009** (0.004)	-0.009** (0.004)	-0.011*** (0.004)	0.009 (0.065)	0.010 (0.030)	0.166 (0.183)	-0.099* (0.053)	-0.070 (0.056)	-0.167** (0.076)	-0.094* (0.056)
Speeches by others	-0.006** (0.003)	-0.004 (0.002)	-0.012** (0.005)	0.000 (0.003)	-0.000 (0.003)	0.003 (0.003)	0.000 (0.002)	0.012 (0.029)	0.006 (0.013)	0.082 (0.083)	0.004 (0.035)	0.023 (0.029)	-0.086 (0.058)	0.005 (0.037)
Speeches by president	-0.010*** (0.004)	-0.006** (0.002)	-0.014*** (0.005)	-0.004* (0.002)	-0.002 (0.002)	-0.004 (0.003)	-0.002 (0.002)	-0.016 (0.028)	0.016 (0.014)	-0.070 (0.085)	0.016 (0.029)	0.018 (0.027)	0.010 (0.037)	0.026 (0.030)
Tweet	-0.005* (0.003)	-0.003 (0.002)	-0.012*** (0.004)	-0.003 (0.003)	-0.004* (0.002)	-0.003 (0.003)	-0.003 (0.002)	0.026 (0.025)	0.017 (0.011)	0.046 (0.087)	0.009 (0.033)	0.036 (0.027)	-0.090 (0.057)	0.018 (0.035)
Panel B: Overall response														
Press Conference	-0.089	-0.081	-0.068	-0.030	-0.022	-0.014	-0.020	-0.051	0.059	-0.014	0.069	0.008	-0.022	0.078
Std. error	0.018	0.014	0.027	0.018	0.017	0.020	0.016	0.091	0.037	0.262	0.112	0.090	0.183	0.121
p-value	0.000	0.000	0.010	0.089	0.190	0.484	0.207	0.576	0.115	0.957	0.539	0.929	0.904	0.520
Whatever it takes	-0.043	0.094	-0.101	0.360	0.358	0.339	0.420	-0.059	-0.014	-1.202	0.153	-1.663	-0.531	0.147
Std. error	0.057	0.047	0.085	0.055	0.052	0.067	0.056	0.371	0.188	1.089	0.408	0.486	0.421	0.403
p-value	0.450	0.044	0.236	0.000	0.000	0.000	0.000	0.873	0.942	0.271	0.707	0.001	0.208	0.716
Observations	2,532	2,532	2,532	2,532	2,532	2,527	2,531	1,461	2,522	254	1,177	1,516	679	1,139
R-squared	0.102	0.139	0.074	0.085	0.057	0.074	0.128	0.055	0.045	0.146	0.046	0.032	0.077	0.049

Notes: The table shows coefficient estimates for the effect of ECB communication events on average absolute favourableness, based on Equation (2.6), for different definitions of non-experts and experts. The models control for day of week, month of year and holiday effects and allow for a linear and quadratic time trend. They contain 5 (1) leads and 3 (2) lags for the press conference and 20 (15) lags for “Whatever it takes” in the English (German) sample. Panel B reports the accumulated response coefficients over all leads and lags, with coefficients that are statistically significant at least at the 10% level printed in bold. Numbers in brackets are standard errors. ***/**/* denote statistical significance at the 1%/5%/10% level.

Table B.13. Standard Deviation of Absolute Favourableness, Robustness Tests

Standard deviation of absolute favourableness														
	English sample							German sample						
	Non-experts (non-centr.) (bm)	Non-experts (non-centr.) (few follow.)	Non-experts (few follow.)	Experts (bm)	Experts (0.33)	Experts (0.75)	Experts (centric)	Non-experts (bm)	Non-experts (excl centr.)	Non-experts (few follow.)	Experts (bm)	Experts (0.33)	Experts (0.75)	Experts (ECB centric)
Panel A: Contemporaneous response														
Press Conference	-0.035*** (0.004)	-0.024*** (0.003)	-0.034*** (0.005)	-0.009*** (0.004)	-0.004 (0.004)	-0.010*** (0.003)	-0.003 (0.003)	0.155*** (0.031)	0.055*** (0.012)	0.005 (0.015)	0.230*** (0.027)	0.234*** (0.023)	0.256*** (0.032)	0.212*** (0.028)
Whatever it takes	-0.022*** (0.005)	-0.009** (0.004)	-0.023*** (0.007)	0.014*** (0.004)	0.020*** (0.005)	0.014*** (0.004)	0.026*** (0.004)	0.276*** (0.029)	0.113*** (0.015)	-0.073 (0.049)	0.229*** (0.032)	0.183*** (0.030)	0.220*** (0.042)	0.213*** (0.033)
Economic Bulletin	-0.005 (0.005)	-0.007** (0.004)	-0.005 (0.006)	-0.006 (0.004)	-0.008* (0.005)	-0.007* (0.004)	-0.008* (0.004)	-0.019 (0.031)	-0.004 (0.017)	-0.003 (0.032)	-0.006 (0.036)	-0.013 (0.034)	0.049 (0.054)	-0.010 (0.037)
Accounts	-0.010* (0.006)	-0.009** (0.005)	-0.001 (0.009)	-0.011** (0.004)	-0.007 (0.005)	-0.009** (0.004)	-0.010** (0.005)	0.023 (0.049)	-0.007 (0.020)	-0.021 (0.022)	-0.035 (0.034)	-0.034 (0.035)	-0.035 (0.038)	-0.033 (0.036)
Speeches by others	-0.003 (0.003)	-0.002 (0.002)	-0.008** (0.004)	0.002 (0.003)	0.004 (0.004)	-0.001 (0.003)	0.002 (0.003)	-0.004 (0.016)	0.010 (0.010)	-0.009 (0.011)	0.019 (0.020)	0.047** (0.019)	0.043* (0.023)	0.015 (0.021)
Speeches by president	-0.008** (0.003)	-0.007*** (0.002)	-0.009** (0.005)	-0.008*** (0.003)	-0.007** (0.003)	-0.005** (0.002)	-0.006** (0.002)	0.052** (0.022)	0.027*** (0.009)	0.075* (0.039)	0.133*** (0.020)	0.105*** (0.019)	0.125*** (0.025)	0.143*** (0.020)
Tweet	-0.004 (0.002)	-0.002 (0.002)	-0.010*** (0.004)	-0.002 (0.003)	-0.002 (0.003)	-0.004 (0.002)	-0.003 (0.003)	0.006 (0.014)	0.018** (0.009)	-0.004 (0.011)	0.007 (0.019)	0.033*** (0.017)	0.043** (0.021)	0.005 (0.020)
Panel B: Overall response														
Press Conference	-0.060	-0.078	-0.001	-0.002	0.061	-0.017	-0.004	0.269	0.147	0.083	0.494	0.431	0.371	0.452
Std. error	0.015	0.013	0.024	0.018	0.022	0.017	0.019	0.058	0.028	0.060	0.069	0.057	0.088	0.068
p-value	0.000	0.000	0.962	0.898	0.006	0.317	0.830	0.000	0.000	0.172	0.000	0.000	0.000	0.000
Whatever it takes	-0.058	0.012	0.007	0.247	0.189	0.184	0.220	2.403	1.304	0.628	1.527	2.039	0.800	1.659
Std. error	0.050	0.043	0.077	0.057	0.072	0.053	0.063	0.229	0.152	0.161	0.213	0.271	0.194	0.195
p-value	0.241	0.785	0.924	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	2,532	2,532	2,532	2,532	2,527	2,532	2,531	1,461	2,522	254	1,177	1,516	679	1,139
R-squared	0.079	0.132	0.048	0.090	0.134	0.060	0.117	0.097	0.087	0.282	0.184	0.142	0.250	0.196

Notes: The table shows coefficient estimates for the effect of ECB communication events on the standard deviation of absolute favourableness, based on Equation (2.6), for different definitions of non-experts and experts. The models control for day of week, month of year and holiday effects and allow for a linear and quadratic time trend. They contain 5 (1) leads and 3 (2) lags for the press conference and 20 (15) lags for “Whatever it takes” in the English (German) sample. Panel B reports the accumulated response coefficients over all leads and lags, with coefficients that are statistically significant at least at the 10% level printed in bold. Numbers in brackets are standard errors. ***/**/* denote statistical significance at the 1%/5%/10% level.

Appendix to Chapter 3: Incentivizing Inflation Expectations

C.1 Other Tables and Figures

Table C.1. Sample Comparisons: Across Groups and SCE

	Flat	Prior	Posterior	Both	Full Sample	SCE Sample
<i>Age</i>						
Under 30	18.8	17.2	17.6	14.4	17.0	11.7
30-39	26.8	26.0	28.4	26.8	27.0	19.0
40-49	25.2	24.0	26.8	24.0	25.0	18.8
50-59	15.2	18.8	14.4	18.0	16.6	20.6
60 or over	14.0	14.0	12.8	16.8	14.4	29.9
<i>Gender</i>						
Female	54.8	65.2	50.8	63.6	58.6	48.1
Male	44.8	34.8	49.2	36.0	41.2	51.9
Prefer not to say	0.4			0.4	0.2	
<i>Income</i>						
Less than \$50,000	48.8	43.6	39.2	38.4	42.5	42.8
\$50,000-\$99,999	30.0	34.0	36.4	39.6	35.0	34.5
\$100,000 or more	21.2	22.4	24.4	22.0	22.5	22.7
<i>Race/Ethnicity</i>						
Asian	6.0	7.2	7.6	8.0	7.2	3.5
Black	14.4	13.6	6.4	14.0	12.1	10.4
White	73.2	69.6	73.6	70.8	71.8	81.8
Other	6.4	9.6	12.4	7.2	8.9	4.4

Notes: Each value in this table represents the percentage of the sample belonging to the corresponding category. Survey of Consumer Expectations (SCE) sample values are taken from Armantier et al. (2017).

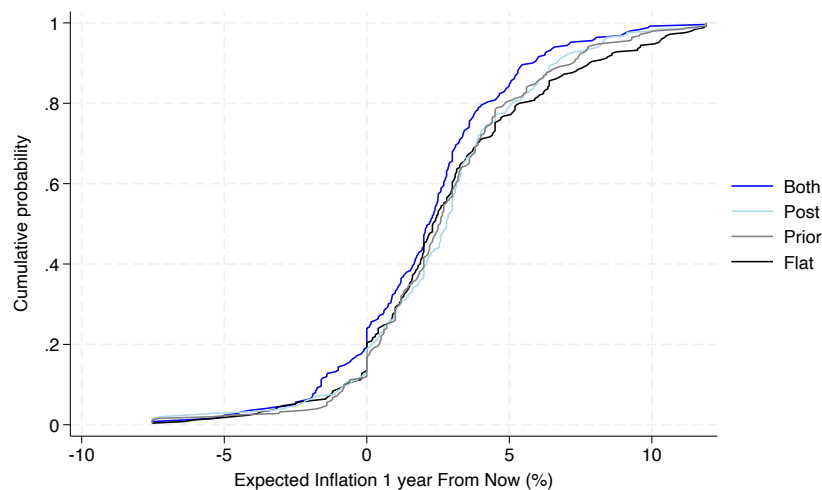
Table C.2. Hypothetical Earnings From Inflation Expectations (The Priors)

	(1)	(2)	(3)	(4)
	Payoff	Payoff	Payoff	Payoff
Post	-0.0606 (0.270)	-0.110 (0.264)	-0.150 (0.264)	-0.197 (0.262)
Both	0.545* (0.281)	0.644** (0.277)	0.626** (0.277)	0.613** (0.275)
Prior	0.755*** (0.282)	0.867*** (0.275)	0.880*** (0.275)	0.874*** (0.274)
Male		1.127*** (0.202)	1.096*** (0.202)	1.112*** (0.202)
Higher_Ed			0.546*** (0.195)	0.537*** (0.195)
Constant	2.296*** (0.193)	1.791*** (0.203)	1.538*** (0.219)	1.795*** (0.268)
Sentiment	No	No	No	Yes
N	1000	1000	1000	1000

Robust standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Notes: This Table shows the effect of treatments on hypothetical earnings, as proxied by distance between reported priors and the Fed's median 2025 forecast. Regressions are estimated by OLS with robust standard errors. Data are winsorized at the 1% and 99% levels.

Figure C.1. CDFs of Inflation Expectations After Information Provision

Notes: The Figure shows cumulative distribution functions (CDFs) of inflation expectations elicited after the information intervention. Expectations are shown by the different treatment groups and expressed in percentage points. Data are winsorized at the 1% and 99% levels.

Table C.3. Summary Statistics and Variance Comparison of APE

	Mean	Std	Median	IQR	N
Unincentivized - Male	10.43	19.18	2.8	6.7	235
Unincentivized - Female	23.48	28.03	9.8	30.0	265
Incentivized - Male	5.71	12.87	1.8	4.0	177
Incentivized - Female	11.52	19.87	2.8	11.7	323
All Data	13.41	22.16	3.2	12.0	1,000

Notes: This table shows mean, median, standard deviation and IQR of the absolute perception error (APE) by gender and incentive treatments. Unincentivized is comprised of treatments *Flat* and *Posterior*, while Incentivized is comprised of *Both* and *Prior*.

Table C.4. Effects of Incentives on Updated Inflation Expectations (The Posteriors)

	(1) $\mathbb{E}(\pi_{Post.})$	(2) $\mathbb{E}(\pi_{Post.})$	(3) $\mathbb{E}(\pi_{Post.})$	(4) $\mathbb{E}(\pi_{Post.})$	(5) $\mathbb{E}(\pi_{Post.})$
Post	-0.162 (0.232)	-0.163 (0.232)	-0.149 (0.231)	-0.115 (0.233)	-0.0885 (0.222)
Both	-0.642*** (0.227)	-0.624*** (0.227)	-0.652*** (0.227)	-0.637*** (0.228)	-0.688*** (0.219)
Prior	-0.339 (0.234)	-0.344 (0.234)	-0.377 (0.234)	-0.387* (0.234)	-0.426* (0.222)
Deflation		-0.316 (0.246)	-0.321 (0.245)	-0.315 (0.245)	-0.0673 (0.243)
Male			-0.325** (0.163)	-0.299* (0.162)	-0.241 (0.157)
Higher_Ed				-0.459*** (0.159)	-0.430*** (0.152)
Constant	4.571**** (0.171)	4.613**** (0.175)	4.759**** (0.192)	4.971**** (0.207)	4.681**** (0.216)
Sentiment	No	No	No	No	Yes
N	1000	1000	1000	1000	1000

Robust standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Notes: This table reports the results of a series of OLS regressions (with robust standard errors) wherein we project inflation expectations estimated using participants' probabilistic inflation forecasts onto a series of dummies denoting treatment and other conditioning information. Data are winsorized at the 1% and 99% levels.

Table C.5. Effect of Incentives on Effort

	(1)	(2)	(3)	(4)
	Completion Time	Completion Time	Completion Time	Completion Time
Post	27.06 (29.01)	30.49 (28.86)	35.13 (28.75)	37.26 (28.86)
Both	127.5**** (30.55)	120.6**** (30.47)	122.7**** (30.26)	119.6**** (30.23)
Prior	58.49** (27.39)	50.69* (27.19)	49.28* (27.19)	53.15* (27.28)
Male		-77.99**** (20.97)	-74.41**** (20.90)	-80.11**** (20.85)
Higher_Ed			-63.14*** (20.55)	-67.40*** (20.54)
Constant	572.3**** (19.99)	607.2**** (21.76)	636.4**** (24.28)	590.7**** (27.78)
Sentiment	No	No	No	Yes
N	1000	1000	1000	1000

Robust standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Notes: This Table shows the effect of treatments on effort, as proxied by completion times. Regressions are estimated by OLS with robust standard errors. Data are winsorized at the 5% and 95% levels.

C.2 Power Analysis

Table C.6. Sample Size Calculation

α	$1 - \beta$	N	N_C	N_T	Δ	μ_C	μ_T	σ
0.01	0.80	1,172	586	586	0.2	0	0.2	1
0.01	0.80	524	262	262	0.3	0	0.3	1
0.01	0.80	296	148	148	0.4	0	0.4	1
0.01	0.80	192	96	96	0.5	0	0.5	1
0.01	0.80	134	67	67	0.6	0	0.6	1
0.01	0.80	100	50	50	0.7	0	0.7	1
0.01	0.80	78	39	39	0.8	0	0.8	1
0.01	0.80	62	31	31	0.9	0	0.9	1
0.01	0.80	52	26	26	1.0	0	1.0	1
0.05	0.80	788	394	394	0.2	0	0.2	1
0.05	0.80	352	176	176	0.3	0	0.3	1
0.05	0.80	200	100	100	0.4	0	0.4	1
0.05	0.80	128	64	64	0.5	0	0.5	1
0.05	0.80	90	45	45	0.6	0	0.6	1
0.05	0.80	68	34	34	0.7	0	0.7	1
0.05	0.80	52	26	26	0.8	0	0.8	1
0.05	0.80	42	21	21	0.9	0	0.9	1
0.05	0.80	34	17	17	1.0	0	1.0	1
0.10	0.80	620	310	310	0.2	0	0.2	1
0.10	0.80	278	139	139	0.3	0	0.3	1
0.10	0.80	156	78	78	0.4	0	0.4	1
0.10	0.80	102	51	51	0.5	0	0.5	1
0.10	0.80	72	36	36	0.6	0	0.6	1
0.10	0.80	52	26	26	0.7	0	0.7	1
0.10	0.80	42	21	21	0.8	0	0.8	1
0.10	0.80	32	16	16	0.9	0	0.9	1
0.10	0.80	28	14	14	1.0	0	1.0	1

Notes: Results are sorted by α and Cohen's d (i.e. μ_T).

Here we determine the necessary sample size for detecting effects of various magnitudes with a significance level of 0.05 and a power of 0.80. Effect magnitudes are specified in terms of Cohen's d, ranging from 0.2 to 1.0 in increments of 0.1. The effect magnitude (Cohen's d) is calculated as the standardized mean difference between the treatment and control groups. Specifically, Cohen's d is defined as:

$$d = \frac{M_1 - M_2}{SD_{pooled}}$$

where M_1 and M_2 are the means of the treatment and control groups, respectively, and SD_{pooled} is the pooled standard deviation of the two groups. We assume $M_1 = 0$, treating it as the control group.

Conventional thresholds for interpreting the magnitude of effect sizes:

- **Small effect size:** $d = 0.2$

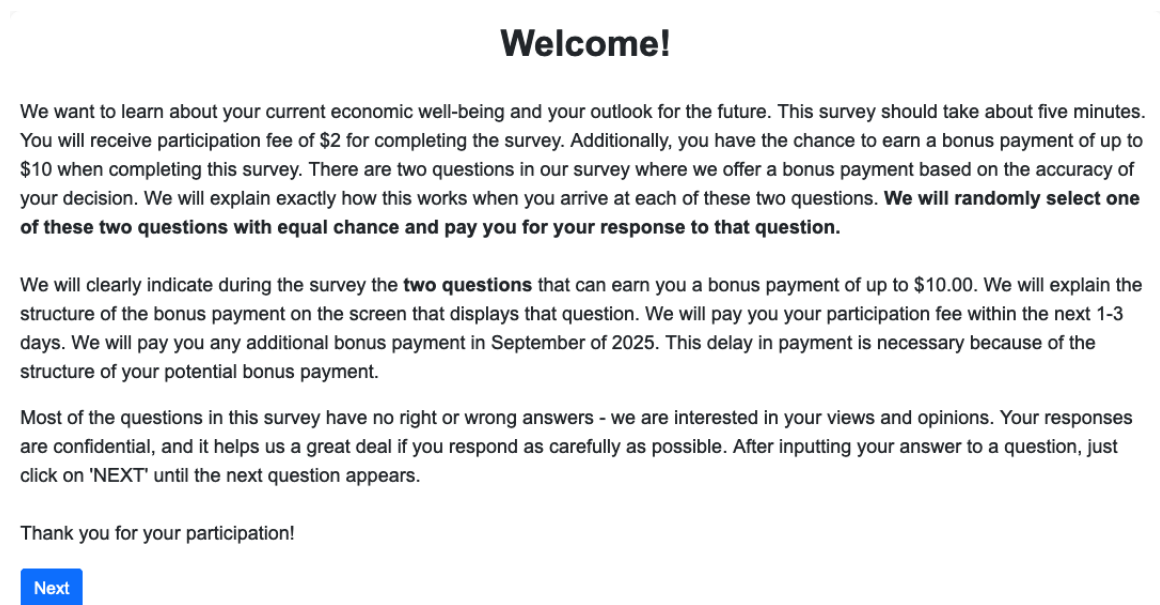
- **Medium effect size:** $d = 0.5$
- **Large effect size:** $d = 0.8$

We base our sample size on this ex-ante power calculation. Our desire to precisely estimate null effects led us to choose a sample size of 250 subjects per treatment. This would allow us to detect small differences via pair-wise comparisons at a one-percent level of significance and $\beta = .8$.

C.3 Inflation Expectations Survey

This section presents the full survey used in this study, which elicits inflation expectations and implements an information provision intervention.

Figure C.2. Welcome



Notes: This figure shows the welcome page for the *Both* treatment group. Slight variations in wording occur between treatments to reflect the different incentive structures. Screenshots of other treatment groups are available upon request.

Figure C.3. General Questions

Please answer the following questions about your financial well-being:

Do you think you (and any family living with you) are financially better or worse off these days than you were twelve months ago?

And looking ahead, do you think you (and any family living with you) will be financially better or worse off twelve months from now than you are these days?

Looking ahead, do you think the economy in the United States will be stronger or weaker twelve months from now than these days?

Next

Figure C.4. Explanations

Next, we would like to ask you for your expectations about the economy. Of course, no one can know the future. These questions have no right or wrong answers - we are interested in your views and opinions.

In some of the following questions, we will ask you to think about the percent chance of something happening in the future. Your answers can range from 0 to 100, where 0 means there is absolutely no chance, and 100 means that it is absolutely certain.

For example, numbers like:

2 and 5 percent may indicate "almost no chance"

18 percent or so may mean "not much chance"

47 or 52 percent chance may be a "pretty even chance"

83 percent or so may mean a "very good chance"

95 or 98 percent chance may be "almost certain"

Next

Figure C.5. Inflation Point Forecast (Flat)

Inflation

This question asks you to forecast inflation. We will use your forecast to compare to the Personal Consumption Expenditures (PCE) inflation for the U.S., which is the Federal Reserve's preferred measure of inflation. This is typically the measure of inflation the Fed discusses and, consequently, you hear discussed publicly.

Over the past twelve months...

Do you think that there was **inflation or deflation**?

And how much **inflation/deflation** do you think there was?

Over the next twelve months...

Do you think that there will be **inflation or deflation**?

And how much **inflation/deflation** do you expect?

Next

Figure C.6. Inflation Point Forecast (Posterior)

Inflation

This question asks you to forecast inflation. We will use your forecast to compare to the Personal Consumption Expenditures (PCE) inflation for the U.S., which is the Federal Reserve's preferred measure of inflation. This is typically the measure of inflation the Fed discusses and, consequently, you hear discussed publicly. the U.S. Bureau of Economic Analysis (BEA) will release information on the most updated measure of Personal Consumption Expenditures (PCE) inflation for the U.S., which is the Federal Reserve's preferred measure of inflation.

Over the past twelve months...

Do you think that there was **inflation or deflation**?

And how much **inflation/deflation** do you think there was?

Over the next twelve months...

Do you think that there will be **inflation or deflation**?

And how much **inflation/deflation** do you expect?

[Next](#)

Figure C.7. Inflation Point Forecast (Prior)

Inflation

Bonus Payment: YOU CAN RECEIVE A BONUS PAYMENT FOR THIS QUESTION. THIS IS THE ONLY QUESTION IN OUR SURVEY FOR WHICH YOU CAN RECEIVE A BONUS PAYMENT.

You can earn up to \$10 for this task. This is in addition to your participation fee. In September of 2025, the U.S. Bureau of Economic Analysis (BEA) will release information on the most updated measure of Personal Consumption Expenditures (PCE) inflation for the U.S., which is the Federal Reserve's preferred measure of inflation.

Once the BEA publishes the actual PCE inflation reported in 12 months, we will compare your forecast to it and pay you based on the accuracy of your forecast. Your bonus payment halves each time your forecast increases by 1 percentage point.

For example:

If your forecast matches the inflation rate exactly, you will earn \$10.

If your forecast is 1 percentage point above or below the inflation rate, you will earn \$5.

If your forecast is 2 percentage points above or below the inflation rate, you will earn \$2.5.

Over the past twelve months...

Do you think that there was **inflation or deflation**?

And how much **inflation/deflation** do you think there was?

Over the next twelve months...

Do you think that there will be **inflation or deflation**?

And how much **inflation/deflation** do you expect?

Next

Figure C.8. Inflation Point Forecast (Both)

Inflation

Bonus Payment: YOU MAY RECEIVE A BONUS PAYMENT FOR THIS QUESTION. THIS IS ONE OF THE TWO QUESTIONS IN OUR SURVEY FOR WHICH YOU CAN RECEIVE A BONUS PAYMENT.

If randomly selected for payment, you can earn up to \$10 for this task. This is in addition to your participation fee. In September of 2025, the U.S. Bureau of Economic Analysis (BEA) will release information on the most updated measure of Personal Consumption Expenditures (PCE) inflation for the U.S., which is the Federal Reserve's preferred measure of inflation.

Once the BEA publishes the actual PCE inflation reported in 12 months, we will compare your forecast to it and pay you based on the accuracy of your forecast. Your bonus payment halves each time your forecast increases by 1 percentage point.

For example:

If your forecast matches the inflation rate exactly, you will earn \$10.

If your forecast is 1 percentage point above or below the inflation rate, you will earn \$5.

If your forecast is 2 percentage points above or below the inflation rate, you will earn \$2.5.

Over the past twelve months...

Do you think that there was **inflation or deflation**?

And how much **inflation/deflation** do you think there was?

Over the next twelve months...

Do you think that there will be **inflation or deflation**?

And how much **inflation/deflation** do you expect?

Next

Figure C.9. Food Point Forecast

Food Prices

Over the past twelve months...

Do you think the **price of food** has increased or decreased?

And by about what percentage do you think the **price of food** has changed?

Over the next twelve months...

Do you think the **price of food** will have increased or decreased?

And by about what percentage do you think the **price of food** will have changed?

Next

Figure C.10. Gas Point Forecast

Gas Prices

Over the past twelve months...

Do you think the **price of a gallon of gas** has increased or decreased?

And by about what percentage do you think the **price of a gallon of gas** has changed?

Over the next twelve months...

Do you think the **price of a gallon of gas** will have increased or decreased?

And by about what percentage do you think the **price of a gallon of gas** will have changed?

Next

Figure C.11. Information Intervention

We provide below the most recent official economic forecast data from the Federal Reserve, which is the central bank for the United States. Economic forecasts are very important for the Fed because policymakers there use forecasts to help them make good policy decisions when guiding our economy.

In conjunction with the Federal Open Market Committee (FOMC) meeting held on June 11–12, 2024, meeting participants submitted their projections of the most likely outcomes for inflation for each year from 2024 to 2026 and over the longer run. We have summarized these projections in the following table:

Variable	Median 2024	Range 2024	Median 2025	Range 2025
PCE inflation	2.6	2.5–3.0	2.3	2.2–2.5

Next

Figure C.12. Food Bin Forecast

Food Prices

Now we would like you to think about the different things that may happen to **food prices** over the **next twelve months**. We realize that this question may take a little more effort.

Below, we will ask you to assign a percent (%) chance that food prices twelve months from now will fall into a certain range. The sum of the numbers you enter should equal 100%. For example, if you think there is a 20% chance that food prices will be 12% or higher, and a 30% chance that it will be between 8% and 12%, you should indicate this by entering the values 20 and 30 into each corresponding field. In this scenario, you would need to allocate the remaining 50%.

In your view, what would you say is the percent chance that, **over the next twelve months**...

the price of food will increase by 12% or higher:		%
the price of food will increase by between 8% and 12%:		%
the price of food will increase by between 4% and 8%:		%
the price of food will increase by between 2% and 4%:		%
the price of food will increase by between 0% and 2%:		%
the price of food will decrease by between -2% and 0%:		%
the price of food will decrease by between -4% and -2%:		%
the price of food will decrease by between -8% and -4%:		%
the price of food will decrease by between -12% and -8%:		%
the price of food will decrease by -12% or lower:		%

Next

Figure C.13. Gas Bin Forecast

Gas Prices

Now we would like you to think about the different things that may happen to **gas prices** over the **next twelve months**. We realize that this question may take a little more effort.

Below, we will ask you to assign a percent (%) chance that gas prices twelve months from now will fall into a certain range. The sum of the numbers you enter should equal 100%. For example, if you think there is a 20% chance that gas prices will be 12% or higher, and a 30% chance that it will be between 8% and 12%, you should indicate this by entering the values 20 and 30 into each corresponding field. In this scenario, you would need to allocate the remaining 50%.

In your view, what would you say is the percent chance that, **over the next twelve months...**

the price of a gallon of gas will increase by 12% or higher:	%
the price of a gallon of gas will increase by between 8% and 12%:	%
the price of a gallon of gas will increase by between 4% and 8%:	%
the price of a gallon of gas will increase by between 2% and 4%:	%
the price of a gallon of gas will increase by between 0% and 2%:	%
the price of a gallon of gas will decrease by between -2% and 0%:	%
the price of a gallon of gas will decrease by between -4% and -2%:	%
the price of a gallon of gas will decrease by between -8% and -4%:	%
the price of a gallon of gas will decrease by between -12% and -8%:	%
the price of a gallon of gas will decrease by -12% or lower:	%

Next

Figure C.14. Inflation Bin Forecast (Flat and Prior)

Inflation

Now we would like you to think about the different things that may happen to **inflation** over the **next twelve months**. We realize that this question may take a little more effort.

Below, we will ask you to assign a percent (%) chance that inflation months from now will fall into a certain range. The sum of the numbers you enter should equal 100%. For example, if you think there is a 20% chance that inflation will be 12% or higher, and a 30% chance that it will be between 8% and 12%, you should indicate this by entering the values 20 and 30 into each corresponding field. In this scenario, you would need to allocate the remaining 50%.

In your view, what would you say is the percent chance that, **over the next twelve months**...

the rate of inflation will be 12% or higher:	%
the rate of inflation will be between 8% and 12%:	%
the rate of inflation will be between 4% and 8%:	%
the rate of inflation will be between 2% and 4%:	%
the rate of inflation will be between 0% and 2%:	%
the rate of deflation (opposite of inflation) will be between 0% and 2%:	%
the rate of deflation (opposite of inflation) will be between 2% and 4%:	%
the rate of deflation (opposite of inflation) will be between 4% and 8%:	%
the rate of deflation (opposite of inflation) will be between 8% and 12%:	%
the rate of deflation (opposite of inflation) will be 12% or lower:	%

Next

Figure C.15. Inflation Bin Forecast (Posterior)

Inflation

Now we would like you to think about the different things that may happen to **inflation** over the **next twelve months**. We realize that this question may take a little more effort.

Below, we will ask you to assign a percent (%) chance that inflation months from now will fall into a certain range. The sum of the numbers you enter should equal 100%. For example, if you think there is a 20% chance that inflation will be 12% or higher, and a 30% chance that it will be between 8% and 12%, you should indicate this by entering the values 20 and 30 into each corresponding field. In this scenario, you would need to allocate the remaining 50%.

Bonus Payment: YOU CAN RECEIVE A BONUS PAYMENT FOR THIS QUESTION. THIS IS THE ONLY QUESTION IN OUR SURVEY FOR WHICH YOU CAN RECEIVE A BONUS PAYMENT.

You can earn up to \$10 for this task. This is in addition to your participation fee. In September of 2025, the U.S. Bureau of Economic Analysis (BEA) will release information on the most updated measure of Personal Consumption Expenditures (PCE) inflation for the U.S., which is the Federal Reserve's preferred measure of inflation. This value of inflation will fall into one of the bins you see here. Your bonus payment will be \$10 multiplied by the weight (i.e. % chance) you assigned to that bin. For example:

- If you assign a 10% chance to a bin and the actual inflation falls into that bin, you will earn $\$10 * 0.10 = \$1.00$.
- If you assign a 25% chance to a bin and the actual inflation falls into that bin, you will earn $\$10 * 0.25 = \$2.50$.
- If you assign a 90% chance to a bin and the actual inflation falls into that bin, you will earn $\$10 * 0.9 = \$9.00$.

In your view, what would you say is the percent chance that, **over the next twelve months...**

the rate of inflation will be 12% or higher:	%
the rate of inflation will be between 8% and 12%:	%
the rate of inflation will be between 4% and 8%:	%
the rate of inflation will be between 2% and 4%:	%
the rate of inflation will be between 0% and 2%:	%
the rate of deflation (opposite of inflation) will be between 0% and 2%:	%
the rate of deflation (opposite of inflation) will be between 2% and 4%:	%
the rate of deflation (opposite of inflation) will be between 4% and 8%:	%
the rate of deflation (opposite of inflation) will be between 8% and 12%:	%
the rate of deflation (opposite of inflation) will be 12% or lower:	%

Next

Figure C.16. Inflation Bin Forecast (Both)

Inflation

Now we would like you to think about the different things that may happen to **inflation** over the **next twelve months**. We realize that this question may take a little more effort.

Below, we will ask you to assign a percent (%) chance that inflation months from now will fall into a certain range. The sum of the numbers you enter should equal 100%. For example, if you think there is a 20% chance that inflation will be 12% or higher, and a 30% chance that it will be between 8% and 12%, you should indicate this by entering the values 20 and 30 into each corresponding field. In this scenario, you would need to allocate the remaining 50%.

Bonus Payment: YOU MAY RECEIVE A BONUS PAYMENT FOR THIS QUESTION. THIS IS ONE OF THE TWO QUESTIONS IN OUR SURVEY FOR WHICH YOU CAN RECEIVE A BONUS PAYMENT.

If randomly selected for payment, you can earn up to \$10 for this task. This is in addition to your participation fee. In September of 2025, the U.S. Bureau of Economic Analysis (BEA) will release information on the most updated measure of Personal Consumption Expenditures (PCE) inflation for the U.S., which is the Federal Reserve's preferred measure of inflation. This value of inflation will fall into one of the bins you see here. Your bonus payment will be \$10 multiplied by the weight (i.e. % chance) you assigned to that bin. For example:

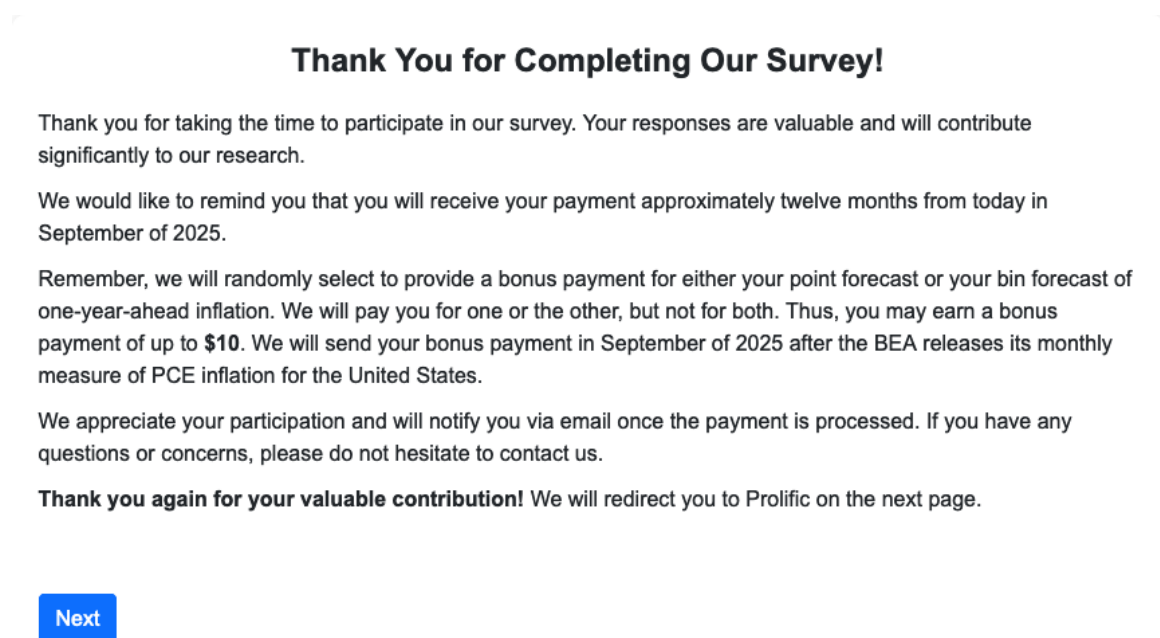
- If you assign a 10% chance to a bin and the actual inflation falls into that bin, you will earn $\$10 \times 0.10 = \$1.00$.
- If you assign a 25% chance to a bin and the actual inflation falls into that bin, you will earn $\$10 \times 0.25 = \$2.50$.
- If you assign a 90% chance to a bin and the actual inflation falls into that bin, you will earn $\$10 \times 0.9 = \$9.00$.

In your view, what would you say is the percent chance that, **over the next twelve months...**

the rate of inflation will be 12% or higher:		%
the rate of inflation will be between 8% and 12%:		%
the rate of inflation will be between 4% and 8%:		%
the rate of inflation will be between 2% and 4%:		%
the rate of inflation will be between 0% and 2%:		%
the rate of deflation (opposite of inflation) will be between 0% and 2%:		%
the rate of deflation (opposite of inflation) will be between 2% and 4%:		%
the rate of deflation (opposite of inflation) will be between 4% and 8%:		%
the rate of deflation (opposite of inflation) will be between 8% and 12%:		%
the rate of deflation (opposite of inflation) will be 12% or lower:		%

Next

Figure C.17. End of Survey



Notes: This figure shows the final page for the *Both* treatment group. Slight variations in wording occur between treatments to reflect the different incentive structures. Screenshots of other treatment groups are available upon request.

Bibliography

- Abeler, J., Huffman, D. B., and Raymond, C. (2023). Incentive complexity, bounded rationality and effort provision. *CESifo Working Paper*.
- Alsan, M., Garrick, O., and Graziani, G. (2019). Does diversity matter for health? experimental evidence from oakland. *American Economic Review*, 109(12):4071–4111.
- Amador, M. and Weill, P.-O. (2010). Learning from prices: Public communication and welfare. *Journal of Political Economy*, 118(5):866–907.
- Andre, P., Pizzinelli, C., Roth, C., and Wohlfart, J. (2022). Subjective models of the macroeconomy: Evidence from experts and representative samples. *The Review of Economic Studies*, 89(6):2958–2991.
- Angeletos, G.-M. and Huo, Z. (2021). Myopia and anchoring. *American Economic Review*, 111(4):1166–1200.
- Angeletos, G.-M. and Lian, C. (2018). Forward guidance without common knowledge. *American Economic Review*, 108(9):2477–2512.
- Angeletos, G.-M. and Pavan, A. (2007). Efficient use of information and social value of information. *Econometrica*, 75(4):1103–1142.
- Angino, S. and Secola, S. (2022). *Instinctive versus reflective trust in the European Central Bank*. Number 2660. ECB Working Paper.
- Armantier, O., Bruine de Bruin, W., Topa, G., Van Der Klaauw, W., and Zafar, B. (2015). Inflation expectations and behavior: Do survey respondents act on their beliefs? *International Economic Review*, 56(2):505–536.
- Armantier, O., Koşar, G., Topa, G., and Van der Klaauw, W. (2024). The survey of consumer expectations: A look back at the past decade. *Liberty Street Economics*.
- Armantier, O., Topa, G., Van der Klaauw, W., and Zafar, B. (2017). An overview of the survey of consumer expectations. *Economic Policy Review*, (23-2):51–72.
- Armantier, O. and Treich, N. (2013). Eliciting beliefs: Proper scoring rules, incentives, stakes and hedging. *European Economic Review*, 62:17–40.
- Azar, P. D. and Lo, A. W. (2016). The wisdom of twitter crowds: Predicting stock market reactions to fomc meetings via twitter feeds. *Institutional Investor Journals*.

REFERENCES

- Bartoš, V., Bauer, M., Chytilová, J., and Matějka, F. (2016). Attention discrimination: Theory and field experiments with monitoring information acquisition. *American Economic Review*, 106(6):1437–1475.
- Bassetto, M. (2019). Forward guidance: Communication, commitment, or both? *Journal of Monetary Economics*, 108:69–86.
- Becker, C. K., Duersch, P., and Eife, T. (2023). Measuring inflation expectations: How the response scale shapes density forecasts. *Available at SSRN 4323706*.
- Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics: Applications and Foundations 1*, 2:69–186.
- Benjamin, D. J., Rabin, M., and Raymond, C. (2016). A model of nonbelief in the law of large numbers. *Journal of the European Economic Association*, 14(2):515–544.
- Bennani, H., Fanta, N., Gertler, P., and Horvath, R. (2020). Does central bank communication signal future monetary policy in a (post)-crisis era? the case of the ecb. *Journal of International Money and Finance*, 104:102167.
- Bergbauer, S., Hernborg, N., Jamet, J.-F., and Persson, E. (2020). The reputation of the euro and the european central bank: interlinked or disconnected? *Journal of European Public Policy*, 27(8):1178–1194.
- Berger, H., Ehrmann, M., and Fratzscher, M. (2011). Monetary policy in the media. *Journal of Money, Credit and Banking*, 43(4):689–709.
- Bholat, D., Broughton, N., Ter Meer, J., and Walczak, E. (2019). Enhancing central bank communications using simple and relatable information. *Journal of Monetary Economics*, 108:1–15.
- Bianchi, F., Gómez-Cram, R., Kind, T., and Kung, H. (2023). Threats to central bank independence: High-frequency identification with twitter. *Journal of Monetary Economics*, 135:37–54.
- Binder, C. (2017). Federal reserve communication and the media. *Journal of Media Economics*, 30(4):191–214.
- Binder, C. (2021). Presidential antagonism and central bank credibility. *Economics & Politics*, 33(2):244–263.
- Blinder, A., Ehrmann, M., De Haan, J., and Jansen, D.-J. (2017). Necessity as the mother of invention: Monetary policy after the crisis. *Economic Policy*, 32(92):707–755.
- Blinder, A. S. (2007). Monetary policy by committee: Why and how? *European Journal of Political Economy*, 23(1):106–123.
- Blinder, A. S. (2018). Through a crystal ball darkly: The future of monetary policy communication. In *AEA Papers and Proceedings*, volume 108, pages 567–571. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.
- Blinder, A. S., Ehrmann, M., De Haan, J., and Jansen, D.-J. (2024). Central bank communication with the general public: Promise or false hope? *Journal of Economic Literature*, 62(2):425–457.

REFERENCES

- Blinder, A. S., Ehrmann, M., Fratzscher, M., De Haan, J., and Jansen, D.-J. (2008). Central bank communication and monetary policy: A survey of theory and evidence. *Journal of Economic Literature*, 46(4):910–945.
- Bodea, C., Ferrara, F. M., Kerner, A., and Sattler, T. (2021). Gender and economic policy: When do women speak with authority on economic issues? evidence from the euro area. *Evidence from the Euro Area* (July 2, 2021).
- Bodea, C. and Kerner, A. (2022). Fear of inflation and gender representation in central banking. *European Journal of Political Economy*, 74:102192.
- Bracha, A. and Tang, J. (2024). Inflation levels and (in) attention. *Review of Economic Studies*.
- Braitsch, H. and Mitchell, J. (2022). A new measure of consumers’(in) attention to inflation. *Economic Commentary*, (2022-14).
- Bruine de Bruin, W., Potter, S., Rich, R. W., Topa, G., and Van der Klaauw, W. (2010). Improving survey measures of household inflation expectations. *Current Issues in Economics and Finance*, 16(7).
- Camerer, C. F. and Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of risk and uncertainty*, 19:7–42.
- Camous, A. and Matveev, D. (2021). Furor over the fed: A president’s tweets and central bank independence. *CESifo Economic Studies*, 67(1):106–127.
- Carrell, S. E., Page, M. E., and West, J. E. (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly journal of economics*, 125(3):1101–1144.
- Cavallo, A., Cruces, G., and Perez-Truglia, R. (2017). Inflation expectations, learning, and supermarket prices: Evidence from survey experiments. *American Economic Journal: Macroeconomics*, 9(3):1–35.
- Charness, G., Gneezy, U., and Halladay, B. (2016). Experimental methods: Pay one or pay all. *Journal of Economic Behavior & Organization*, 131:141–150.
- Chen, D. L., Schonger, M., and Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.
- Coenen, G., Ehrmann, M., Gaballo, G., Hoffmann, P., Nakov, A., Nardelli, S., Persson, E., and Strasser, G. (2017). *Communication of monetary policy in unconventional times*. Number 2080. ECB working paper.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.
- Coibion, O., Georgarakos, D., Gorodnichenko, Y., and Weber, M. (2023). Forward guidance and household expectations. *Journal of the European Economic Association*, 21(5):2131–2171.
- Coibion, O. and Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, 105(8):2644–2678.
- Coibion, O., Gorodnichenko, Y., and Kamdar, R. (2018). The formation of expectations, inflation, and the phillips curve. *Journal of Economic Literature*, 56(4):1447–1491.

REFERENCES

- Coibion, O., Gorodnichenko, Y., Kumar, S., and Pedemonte, M. (2020). Inflation expectations as a policy tool? *Journal of International Economics*, 124:103297.
- Coibion, O., Gorodnichenko, Y., Kumar, S., and Ryngaert, J. (2021). Do you know that i know that you know...? higher-order beliefs in survey data. *The Quarterly Journal of Economics*, 136(3):1387–1446.
- Coibion, O., Gorodnichenko, Y., and Weber, M. (2022). Monetary policy communications and their effects on household inflation expectations. *Journal of Political Economy*, 130(6):1537–1584.
- Conrad, C., Enders, Z., and Glas, A. (2022). The role of information and experience for households' inflation expectations. *European Economic Review*, 143:104015.
- Conti-Brown, P. and Feinstein, B. D. (2020). Twitter and the federal reserve. *Brookings Center on Regulation & Markets Working Paper*.
- Cornand, C. and Heinemann, F. (2008). Optimal degree of public information dissemination. *The Economic Journal*, 118(528):718–742.
- D'Acunto, F., Fuster, A., and Weber, M. (2022). A diverse fed can reach underrepresented groups. *University of Chicago, Becker Friedman Institute for Economics Working Paper*.
- D'Acunto, F., Malmendier, U., and Weber, M. (2023). What do the data tell us about inflation expectations? In *Handbook of economic expectations*, pages 133–161. Elsevier.
- D'Acunto, F. and Weber, M. (2024). Why survey-based subjective expectations are meaningful and important. *Annual Review of Economics*, 16.
- Danilak, M. (2015). langdetect.
- Danz, D., Vesterlund, L., and Wilson, A. J. (2022). Belief elicitation and behavioral incentive compatibility. *American Economic Review*, 112(9):2851–2883.
- De Fiore, F., Lombardi, M. J., and Schuffels, J. (2022). Are households indifferent to monetary policy announcements? *CEPR Discussion Paper No. DP17041*.
- Do Hwang, I., Lustenberger, T., and Rossi, E. (2021). Does communication influence executives' opinion of central bank policy? *Journal of International Money and Finance*, 115:102393.
- Drobot, S., Puzzello, D., Rholes, R., and Wabitsch, A. (2024). Incentivizing inflation expectations. *Available at SSRN 5226305*.
- Ehrmann, M., Soudan, M., and Stracca, L. (2013). Explaining european union citizens' trust in the european central bank in normal and crisis times. *The Scandinavian Journal of Economics*, 115(3):781–807.
- Ehrmann, M. and Wabitsch, A. (2022). Central bank communication with non-experts—a road to nowhere? *Journal of Monetary Economics*, 127:69–85.
- Esponda, I., Oprea, R., and Yuksel, S. (2023). Seeing what is representative. *The Quarterly Journal of Economics*, page qjad020.

REFERENCES

- European Central Bank (2016). What is the 'quiet period'? URL: <https://www.ecb.europa.eu/ecb/educational/explainers/tell-me/html/what-is-the-quiet-period.en.html>. Accessed: June 28, 2023.
- Gächter, S. and Renner, E. (2010). Incentive-compatible belief elicitation and communication. *American Economic Review*, 100(3):835–860.
- Gershenson, S., Hart, C. M., Hyman, J., Lindsay, C. A., and Papageorge, N. W. (2022). The long-run impacts of same-race teachers. *American Economic Journal: Economic Policy*, 14(4):300–342.
- Gnan, P. and Rieder, K. (2023). Who talks during monetary policy quiet periods, and why? evidence from the european central bank's governing council. *Journal article*, 2023.
- Gorodnichenko, Y., Pham, T., and Talavera, O. (2021). Social media, sentiment and public opinions: Evidence from# brexit and# uselection. *European Economic Review*, 136:103772.
- Grewenig, E., Lergetporer, P., Werner, K., and Woessmann, L. (2022). Incentives, search engines, and the elicitation of subjective beliefs: Evidence from representative online survey experiments. *Journal of Econometrics*, 231(1):304–326.
- Gust, C. and Lopez-Salido, D. (2024). Optimal monetary policy with uncertain private sector foresight.
- Haaland, I., Roth, C., and Wohlfart, J. (2023). Designing information provision experiments. *Journal of Economic Literature*, 61(1):3–40.
- Haldane, A., Macaulay, A., and McMahon, M. (2020). The 3 e's of central bank communication with the public.
- Haldane, A. and McMahon, M. (2018). Central bank communications and the general public. In *AEA papers and proceedings*, volume 108, pages 578–583. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.
- Hellwig, C. (2005). Heterogeneous information and the welfare effects of public information disclosures. Technical report, UCLA mimeo.
- Henrique, J. (2016). Getoldtweets-python. *GitHub: San Francisco, CA, USA*, 2019.
- Inoue, A., Kilian, L., and Kiraz, F. B. (2009). Do actions speak louder than words? household expectations of inflation based on micro consumption data. *Journal of Money, Credit and Banking*, 41(7):1331–1363.
- Issing, O. (2019). *The long journey of central bank communication*. MIT Press.
- James, J. G. and Lawler, P. (2011). Optimal policy intervention and the social value of public information. *American Economic Review*, 101(4):1561–1574.
- Keane, M. P. and Runkle, D. E. (1990). Testing the rationality of price forecasts: New evidence from panel data. *The American Economic Review*, pages 714–735.
- Killer, M. (2015). Textblob-de.
- Knotek, E. S., Mitchell, J., Pedemonte, M., and Shiroff, T. (2024). The effects of interest rate

REFERENCES

- increases on consumers' inflation expectations: The roles of informedness and compliance. *FRB of Cleveland Working Paper*.
- Kohlhas, A. N. (2020). An informational rationale for action over disclosure. *Journal of Economic Theory*, 187:105023.
- Kohlhas, A. N. (2022). Learning by sharing: Monetary policy and common knowledge. *American Economic Journal: Macroeconomics*, 14(3):324–364.
- Korhonen, I. and Newby, E. (2019). Mastering central bank communication challenges via twitter. Technical report, BoF Economics Review.
- Kryvtsov, O. and Petersen, L. (2021). Central bank communication that works: Lessons from lab experiments. *Journal of Monetary Economics*, 117:760–780.
- Kumar, S., Afrouzi, H., Coibion, O., and Gorodnichenko, Y. (2015). Inflation targeting does not anchor inflation expectations: Evidence from firms in new zealand. *Brookings papers on economic activity*, 2015(2):151–225.
- Lagarde, C. (2019). Opening statement to the economic and monetary affairs committee of the european parliament, 4 september 2019.
- Lamla, M. and Vinogradov, D. (2022). Is the word of a gentleman as good as his tweet? policy communications of the bank of england.
- Lamla, M. J. and Vinogradov, D. V. (2019). Central bank announcements: Big news for little people? *Journal of Monetary Economics*, 108:21–38.
- Lan, M., Liu, L., Hernandez, A., Liu, W., Zhou, H., and Wang, Z. (2019). The spillover effect of geotagged tweets as a measure of ambient population for theft crime. *Sustainability*, 11(23):6748.
- Lewis, D. J., Makridis, C., and Mertens, K. (2019). Do monetary policy announcements shift household expectations? *FRB of Dallas Working Paper*.
- Loria, S. (2018). textblob. *Release 0.15*, 2:269.
- Loria, S., Keen, P., Honnibal, M., Yankovsky, R., Karesh, D., Dempsey, E., et al. (2014). Textblob: simplified text processing. *Secondary TextBlob: simplified text processing*, 3:2014.
- Lucca, D. O. and Moench, E. (2015). The pre-fomc announcement drift. *The Journal of finance*, 70(1):329–371.
- Lüdering, J. and Tillmann, P. (2020). Monetary policy on twitter and asset prices: Evidence from computational text analysis. *The North American Journal of Economics and Finance*, 51:100875.
- Mackowiak, B. and Wiederholt, M. (2024). Rational inattention during an rct. Technical report.
- Malhotra, N. (2008). Completion time and response order effects in web surveys. *Public opinion quarterly*, 72(5):914–934.
- Malmendier, U. and Veldkamp, L. (2022). Information resonance. Technical report, Working Paper.

REFERENCES

- Manski, C. F. (2004). Measuring expectations. *Econometrica*, 72(5):1329–1376.
- Masciandaro, D., Romelli, D., and Rubera, G. (2021). Monetary policy, twitter and financial markets: evidence from social media traffic. Technical report, BAFFI CAREFIN, Centre for Applied Research on International Markets Banking Finance and Regulation, Università Bocconi.
- McMahon, M. and Rholes, R. (2023). Building central bank credibility: The role of forecast performance. *Centre for Economic Policy Research*.
- Meinusch, A. and Tillmann, P. (2017). Quantitative easing and tapering uncertainty: Evidence from twitter. Technical report. Issue 51.
- Mokhtarzadeh, F. and Petersen, L. (2021). Coordinating expectations through central bank projections. *Experimental economics*, 24:883–918.
- Morris, S. and Shin, H. S. (2002). Social value of public information. *American economic review*, 92(5):1521–1534.
- Mullainathan, S. and Shleifer, A. (2005). The market for news. *American economic review*, 95(4):1031–1053.
- Naveed, N., Gottron, T., Kunegis, J., and Alhadi, A. C. (2011). Bad news travel fast: A content-based analysis of interestingness on twitter. In *Proceedings of the 3rd international web science conference*, pages 1–7.
- Nelson, S. and Bessler, W. (1989). Incentive-compatible scoring rules. *Journal of Economic Theory*, 45(2):123–145.
- Nimark, K. P. and Pitschner, S. (2019). News media and delegated information choice. *Journal of Economic Theory*, 181:160–196.
- Palfrey, T. and Wang, Y. (2009). Incentive-compatible scoring rules for belief elicitation. *Econometrica*, 77(6):2153–2177.
- Pesaran, M. H. and Weale, M. (2006). Survey expectations. *Handbook of economic forecasting*, 1:715–776.
- Petersen, L. and Rholes, R. (2022). Macroeconomic expectations, central bank communication, and background uncertainty: A covid-19 laboratory experiment. *Journal of Economic Dynamics and Control*, 143:104460.
- Pfajfar, D. and Žakelj, B. (2016). Uncertainty in forecasting inflation and monetary policy design: Evidence from the laboratory. *International Journal of Forecasting*, 32(3):849–864.
- Phillips, L. D. and Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of experimental psychology*, 72(3):346.
- Poldi, F. and Zacharias, C. (2018). Twint project. <https://github.com/twintproject/twint>.
- Price, J. (2010). The effect of instructor race and gender on student persistence in stem fields. *Economics of Education Review*, 29(6):901–910.

REFERENCES

- Rholes, R. and Petersen, L. (2021). Should central banks communicate uncertainty in their projections? *Journal of Economic Behavior & Organization*, 183:320–341.
- Roth, C. and Wohlfart, J. (2020). How do expectations about the macroeconomy affect personal expectations and behavior? *Review of Economics and Statistics*, 102(4):731–748.
- Schnabel, I. (2020). Narratives about the ecb’s monetary policy – reality or fiction? speech at the juristische studiengesellschaft. https://www.ecb.europa.eu/press/key/date/2020/html/ecb.sp200211_1~b439a2f4a0.en.html.
- Smith, V. L. (1976). Experimental economics: Induced value theory. *The American Economic Review*, 66(2):274–279.
- Smith, V. L. and Walker, J. M. (1993). Monetary rewards and decision cost in experimental economics. *Economic inquiry*, 31(2):245–261.
- Song, S. and Miled, Z. B. (2017). Digital immunization surveillance: monitoring flu vaccination rates using online social networks. In *2017 IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, pages 560–564. IEEE.
- Stantcheva, S. (2023). How to run surveys: A guide to creating your own identifying variation and revealing the invisible. *Annual Review of Economics*, 15(1):205–234.
- Stiefel, M. and Vivès, R. (2022). ‘whatever it takes’ to change belief: evidence from twitter. *Review of World Economics*, 158(3):715–747.
- Stolper, O. and Walter, A. (2019). Birds of a feather: The impact of homophily on the propensity to follow financial advice. *The Review of Financial Studies*, 32(2):524–563.
- Svensson, L. E. (2006). Social value of public information: Morris and shin (2002) is actually pro-transparency, not con. *American Economic Review*, 96(1):448–452.
- Tavazoei, F., Conversano, C., and Mola, F. (2017). Investigating the relationship between tweeting style and popularity: the case of us presidential election 2016. In *Knowledge Engineering and Semantic Web: 8th International Conference, KESW 2017, Szczecin, Poland, November 8-10, 2017, Proceedings 8*, pages 112–123. Springer.
- Ter Ellen, S., Larsen, V. H., and Thorsrud, L. A. (2022). Narrative monetary policy surprises and the media. *Journal of Money, Credit and Banking*, 54(5):1525–1549.
- Tillmann, P. (2020). Trump, twitter, and treasuries. *Contemporary Economic Policy*, 38(3):403–408.
- Trautmann, S. and van de Kuilen, G. (2014). Incentive-compatible scoring rules and applications to prediction markets. *Journal of Economic Theory*, 149:112–134.
- Tutino, A. (2016). Central bank communication must overcome the public’s limited attention span. *Economic Letter*, 11(6):1–4.
- Van der Cruysen, C., Jansen, D.-J., and De Haan, J. (2015). How much does the public know about the ecb’s monetary policy? evidence from a survey of dutch households. *42nd issue (December 2015) of the International Journal of Central Banking*.
- Veldkamp, L. L. (2011). *Information choice in macroeconomics and finance*. Princeton University Press.

REFERENCES

- Wabitsch, A. (2024). The messenger matters. *Available at SSRN* 5238987.
- Wang, Y. (2011). Incentive-compatible scoring rules for informative reporting. *Journal of the European Economic Association*, 9(3):550–569.
- Warner, J. T. and Pleeter, S. (2001). The personal discount rate: Evidence from military downsizing programs. *American Economic Review*, 91(1):33–53.
- Weber, M., Candia, B., Afrouzi, H., Ropele, T., Lluberas, R., Frache, S., Meyer, B., Kumar, S., Gorodnichenko, Y., Georgarakos, D., et al. (2025). Tell me something i don't already know: Learning in low-and high-inflation settings. *Econometrica*, 93(1):229–264.
- Weber, M., D'Acunto, F., Gorodnichenko, Y., and Coibion, O. (2022). The subjective inflation expectations of households and firms: Measurement, determinants, and implications. *Journal of Economic Perspectives*, 36(3):157–184.
- Wojcik, S. and Hughes, A. (2019). Sizing up twitter users. <https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/>.