# Essays on Digital Economics and AI

Dissertation

accepted by the
Faculty of Management, Economics and Social Sciences
at the University of Cologne
for the degree of
Philosophiae Doctor (PhD)

2026

by
Fabian Roeben
from Cologne, Germany

| | |
|---|---|
| First reviewer: | Prof. Christopher Roth, PhD |
| Second reviewer: | Prof. Tom Zimmermann, PhD |
| Date of oral defense: | August 27, 2026 |

# Acknowledgments

While a PhD oftentimes is a solitary endeavor, there is people I want to thank and without whom this dissertation would not have been possible.

First of all, I am deeply grateful to my supervisors, Christopher Roth and Tom Zimmermann, for their guidance, support, and feedback throughout my PhD. I am especially indebted to Chris for his close mentorship, the trust he placed in me, the freedom he gave me to pursue my own curiosity, and ultimately for becoming a good friend. Working with him on research projects throughout my PhD was especially fun and formative; I learned a great deal from our collaboration, from the way he approaches research in creative and ambitious ways, and in general from our conversations throughout the process. His support helped make this dissertation possible and I would not be where I am today without him. I am also grateful to Tom for always being available with his valuable perspective, constructive feedback, and support in sharpening the framing and presentation of this work.

I thank my co-authors for the many discussions, ideas, and joint efforts that shaped the projects included in this dissertation. I learned a great deal from working closely with them, and I am grateful for the collaborative spirit in which these projects developed.

I also thank the University of Cologne and ECONtribute for providing an excellent research environment, financial support, and many opportunities for exchange through seminars, workshops, and other academic events. Presenting work in these settings, and discussing it in more informal conversations, provided many useful comments and helped improve the papers at different stages. I am grateful to the faculty and administrative staff for their support and help with the many practical matters that accompany a PhD.

I am also grateful to my fellow PhD students for discussions about research, practical advice, helpful comments, and for making the day-to-day work during the PhD both intellectually stimulating and enjoyable.

Finally, I thank my family and friends for their unconditional support throughout this process. Their encouragement and patience have been invaluable and gave me stability during the harder phases of my PhD.

# Contents


**Acknowledgments** **i**

**List of Figures** **vi**

**List of Tables** **ix**

**Introduction** **1**

**1 News Customization with AI** **4**

1.1 Introduction . . . . . 5
1.2 The AI News App . . . . . 10
1.2.1 App . . . . . 10
1.2.2 Source of Articles . . . . . 11
1.2.3 Article Versions . . . . . 12
1.2.4 Multi-Agent AI System for Rewrites . . . . . 14
1.2.5 Validation of News Event Consistency . . . . . 16
1.2.6 Human Validation of Rewrites . . . . . 17
1.2.7 Ethical Considerations . . . . . 19
1.3 The Demand for Customization . . . . . 19
1.3.1 Experimental Design . . . . . 19
1.3.2 Results . . . . . 21
1.4 News Consumption With Customization . . . . . 22
1.4.1 Measuring News Preferences with Customization . . . . . 22
1.4.2 Experimental Design . . . . . 23
1.4.3 How Do People Consume News with Customization? . . . . . 28
1.4.4 Does Customization Improve Matching with Preferences? . . . . . 32
1.4.5 Customization, News Satisfaction and App Perceptions . . . . . 36
1.4.6 Customization, Political Polarization, and Factual Knowledge . . . . . 39
1.5 Conclusion . . . . . 41

**2 The Asymmetric Response to Political Spin** **43**

2.1 Introduction . . . . . 44
2.2 The AI News App . . . . . 47
2.2.1 News Sourcing and Feed Curation . . . . . 47
2.2.2 Article Versions . . . . . 48
2.2.3 Rewrite Pipeline and Safeguards . . . . . 49
2.2.4 Validation of Rewrites . . . . . 50
2.2.5 Spin Analysis of Article Versions . . . . . 51

2.3 Experimental Design . . . 53
2.3.1 Overview . . . 53
2.3.2 Recruitment Survey and Sample Selection . . . 54
2.3.3 App Download Survey . . . 55
2.3.4 Surveys . . . 55
2.3.5 Sample . . . 56
2.4 The Effects of Political Spin . . . 57
2.4.1 News Consumption . . . 57
2.4.2 Perceptions and News Satisfaction . . . 59
2.4.3 Knowledge Acquisition . . . 62
2.4.4 Polarization . . . 63
2.4.5 Heterogeneity by Ideology . . . 64
2.5 External Validity . . . 65
2.6 Conclusion . . . 66

**3 Measuring the Structure of Household Macroeconomic Expectations 68**
3.1 Introduction . . . 69
3.2 Related Literature . . . 72
3.3 Conceptual Framework: Why Joint Expectation Movements Matter . . . 73
3.4 Data and Measurement Design . . . 75
3.4.1 SCE data and benchmark sample . . . 75
3.4.2 SCE variables and transformations . . . 76
3.4.3 ECB CES comparison sample . . . 76
3.4.4 Primitive correlations and signed-comovement indices . . . 77
3.5 SCE Measurement Results . . . 78
3.6 SCE Validation . . . 81
3.7 Evidence from the ECB Consumer Expectations Survey . . . 85
3.8 What the Measure Predicts . . . 86
3.8.1 Predicting later expectation dynamics . . . 86
3.8.2 Direct spending and personal-risk prediction . . . 87
3.8.3 Belief-system structure and the inflation–spending slope . . . 88
3.9 Conclusion . . . 90

**Bibliography 92**

**Appendix A: News Customization with AI 103**
A.1 Summary of Data Collections . . . 103
A.2 Validation of News Event Consistency . . . 104
A.3 Human Validation Study . . . 107
A.4 Demand Experiment . . . 108
A.5 Usage Incentive Experiment . . . 113
A.5.1 Sample . . . 113
A.5.2 Measuring Customization Preferences . . . 119

A.5.3 Heterogeneity in News Consumption With Customization . . . . . . . . . 120
A.5.4 Customization and News Consumption . . . . . . . . . . . . . . . . . . . . . . 130
A.5.5 Customization and News Satisfaction . . . . . . . . . . . . . . . . . . . . . . 133
A.5.6 Disaggregated Treatment Effects . . . . . . . . . . . . . . . . . . . . . . . . 135
A.5.7 Meta Study on Effects on Knowledge and Polarization . . . . . . . . . . . 136

**Appendix B: The Asymmetric Response to Political Spin** **137**
B.1 Mobile App: TailorMadeTimes . . . . . . . . . . . . . . . . . . . . . . . . . . 137
B.2 Validation of Facts and Rewrites . . . . . . . . . . . . . . . . . . . . . . . . . 142
B.2.1 Fact Validation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 142
B.2.2 Rewrite Validation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 143
B.3 Classification of Spin . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 144
B.4 Spin Analysis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 146
B.5 Summary Statistics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 149
B.6 Raw Consumption Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 154
B.7 Consumption by Period . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 157
B.8 Exposure Validation During the Incentive Period . . . . . . . . . . . . . . . . 158
B.9 Political Heterogeneity . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 160
B.10 Additional Tables . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 166
B.11 Endline Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 171
B.12 Comparison of Control Conditions . . . . . . . . . . . . . . . . . . . . . . . . 173
B.13 Different Political Alignment Definition . . . . . . . . . . . . . . . . . . . . 175
B.14 User Experience Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 179

**Appendix C: Measuring the Structure of Household Macroeconomic Expectations** **181**
C.1 Survey Instruments and Variable Construction . . . . . . . . . . . . . . . . . 181
C.2 Illustrative Mechanisms for Joint Expectation Movements . . . . . . . . . . . 188
C.3 Reduced-Form Algebra Behind the Slope Prediction . . . . . . . . . . . . . . 188
C.4 Measurement Formula Details . . . . . . . . . . . . . . . . . . . . . . . . . . 189
C.5 Validation Evidence and Interpretation Boundaries . . . . . . . . . . . . . . 191
C.6 Additional SCE Measurement Details . . . . . . . . . . . . . . . . . . . . . . 192
C.7 Sample Robustness . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 197
C.8 Additional SCE Validation . . . . . . . . . . . . . . . . . . . . . . . . . . . . 199
C.9 Additional ECB CES Evidence . . . . . . . . . . . . . . . . . . . . . . . . . . 205
C.10 Exploratory Evidence on Discrete Types . . . . . . . . . . . . . . . . . . . . 207
C.11 Additional Prediction and Behavior Checks . . . . . . . . . . . . . . . . . . . 208

**Appendix D: Experiment Instructions** **215**

**Appendix E: Prompts** **412**

**List of Sources and Aids Used** **456**

**Curriculum Vitae** **457**

**Affidavit** **459**

# List of Figures


1.1 The AI News App . . . 6
1.2 Multi-Agent AI System for Rewrites . . . 15
1.3 Human Validation of Article Rewrites . . . 18
1.4 Demand for News Customization . . . 22
1.5 Overview of the Usage Incentive Experiment . . . 24
1.6 The AI News App: Customization vs. Control Version . . . 26
1.7 News Customization in the Incentive Period . . . 30
1.8 Perceived News Mismatch and Time Spent in Different Customizations . . . 34
1.9 News Customization in the Post-Incentive Period . . . 36
1.10 Effects on News Satisfaction, Perceptions, and Downstream Outcomes . . . 40

2.1 TMT App . . . 49
2.2 Distribution of Spin . . . 52
2.3 Experiment Overview . . . 54
2.4 Effects on News Consumption . . . 60
2.5 Effects on Perceptions and News Satisfaction . . . 61
2.6 Analysis of Open-Ended Responses . . . 62
2.7 Effects on Knowledge and Polarization . . . 64

3.1 Example respondents: four expectations over 12 survey waves . . . 80
3.2 Probability and inflation-link comovement in the SCE . . . 82
3.3 The inflation–spending slope by probability comovement . . . 90
A.1 Human Validation of Article Rewrites by Party Affiliation . . . 107
A.2 Human Validation by Rewrite Dimension . . . 107
A.3 Overview of the Demand Experiment . . . 108
A.4 App Usage Time on Day 1 by Customization Condition . . . 112
A.5 App Usage Frequency and Time by Customization Condition . . . 112
A.6 Sample Size in the Usage Incentive Experiment . . . 113
A.7 Baseline News Consumption . . . 117
A.8 App Usage by Treatment Arm . . . 119
A.9 App Download Rate by Customization . . . 119
A.10 News Customization in the Incentive Period (Unweighted) . . . 120
A.11 News Customization in the Incentive Period by Party Affiliation . . . 120
A.12 News Customization in the Incentive Period by Voting . . . 121
A.13 News Customization in the Incentive Period by Baseline News Consumption . . . 121
A.14 News Customization in the Incentive Period by Baseline News Bias . . . 122
A.15 News Dissatisfaction and Time Spent in Different Customizations . . . 122

A.16 News Sections by Treatment Arm During the Post-Incentive Period . . . . . . . . 124
A.17 News Customization by Section During the Post-Incentive Period . . . . . . . . . 125
A.18 Number of Slider Changes over Time . . . . . . . . . . . . . . . . . . . . . . . . . . 127
A.19 News Customization over Time . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 127
A.20 Trust in AI and Time Spent in Different Customizations . . . . . . . . . . . . . . . 129
A.21 Perceived News Mismatch and Time Spent by Control Arm . . . . . . . . . . . . . 129
A.22 Time Spent During Incentive and No-Incentive Periods by Treatment Arm . . . . 130
A.23 Effect of Customization on Consumption during the Post-Incentive Period . . . . 131
A.24 Disaggregated Effects on Consumption during the Post-Incentive Period . . . . . 131
A.25 News Customization in the Post-Incentive Period . . . . . . . . . . . . . . . . . . . 132
A.26 Effects on News Satisfaction, Perceptions, and Downstream Outcomes at Endline 133
A.27 Disaggregated Effects on News Satisfaction . . . . . . . . . . . . . . . . . . . . . . 134
A.28 News Satisfaction Heterogeneity by Satisfaction at Baseline . . . . . . . . . . . . . 134
A.29 Disaggregated Effects on Perceptions, Knowledge, and Polarization . . . . . . . . 135
A.30 Perceptions by Control Arm . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 135
A.31 Comparison to Other Experimental Estimates of Effects . . . . . . . . . . . . . . . 136
B.1 TMT App: Politically Neutral Versions . . . . . . . . . . . . . . . . . . . . . . . . 137
B.2 Human Validation of Article Rewrites . . . . . . . . . . . . . . . . . . . . . . . . . 144
B.3 Distribution of Spin Across News Categories . . . . . . . . . . . . . . . . . . . . . 146
B.4 Distribution of Spin Within Politically Neutral Versions . . . . . . . . . . . . . . . 147
B.5 Sample size . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 150
B.6 Raw Consumption Data: Time . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 154
B.7 Raw Consumption Data: Days . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 155
B.8 Raw Consumption Data: Articles . . . . . . . . . . . . . . . . . . . . . . . . . . . . 156
B.9 Effects on News Consumption by Period . . . . . . . . . . . . . . . . . . . . . . . . 157
B.10 Effects on News Consumption by Democrats and Republicans . . . . . . . . . . . 159
B.11 Effects on News Consumption by Ideology . . . . . . . . . . . . . . . . . . . . . . 160
B.12 Effects on Perceptions and News Satisfaction by Democrats and Republicans . . . 161
B.13 Effects on Perceptions and News Satisfaction by Ideology . . . . . . . . . . . . . . 161
B.14 Effects of Misaligned Spin on Post-Incentive Consumption by Strength of Partisanship . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 162
B.15 Effects of Misaligned Spin on Perceptions and Satisfaction by Strength of Partisanship 163
B.16 Effects on Knowledge and Polarization by Democrats and Republicans . . . . . . 163
B.17 Effects on Knowledge and Polarization by Ideology . . . . . . . . . . . . . . . . . 164
B.18 Effects on Perceptions of Complexity, Entertainment, Sentiment and Trust in AI . 165
B.19 Effects on Perceptions and News Satisfaction at Endline . . . . . . . . . . . . . . . 171
B.20 Effects on Knowledge and Polarization at Endline . . . . . . . . . . . . . . . . . . 172
B.21 Differences in News Consumption across neutral control conditions . . . . . . . . 173
B.22 Differences in Perceptions and News Satisfaction across neutral control conditions 174
B.23 Differences in Perceptions of Complexity, Entertainment, and Sentiment across neutral control conditions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 175
B.24 Effects on News Consumption (Alternative Alignment Definition) . . . . . . . . . 176

B.25 Effects on Perceptions and News Satisfaction . . . . . 177
B.26 Effects on Knowledge and Polarization . . . . . 178
B.27 Analysis of Open-Ended Responses . . . . . 180
C.1 Primitive SCE within-household correlations . . . . . 192
C.2 SCE signed-comovement index distributions . . . . . 193
C.3 ECB CES adverse macro-comovement distributions . . . . . 205

# List of Tables

1.1 Rewrite Examples . . . 13
1.2 Alignment of Core Facts . . . 17

2.1 Drivers of Spin in Political Rewrites . . . 53

3.1 Within-household correlations between expectations . . . 78
3.2 SCE signed-comovement measures . . . 81
3.3 Reduced SCE measures against within-person time-permutation nulls . . . 82
3.4 Validation hierarchy for SCE comovement measures . . . 83
3.5 ECB CES evidence on adverse macro comovement . . . 86
3.6 First-half comovement and later expectation dynamics . . . 87
3.7 Spending-slope heterogeneity by probability comovement . . . 89
A.1 Overview of Data Collections . . . 103
A.2 Example of Fact Extraction and Classification . . . 105
A.3 Example Facts Alignment . . . 106
A.4 Summary Statistics for the Demand Experiment . . . 109
A.5 Balance Table for the Demand Experiment . . . 110
A.6 App Downloads by Customization Condition . . . 111
A.7 Treatment Arms in the Usage Incentive Experiment . . . 114
A.8 Attrition in the Usage Incentive Experiment . . . 114
A.9 Summary Statistics for the Usage Incentive Experiment . . . 115
A.10 Balance Table for the Usage Incentive Experiment . . . 116
A.11 Selection into Downloading the App . . . 118
A.12 Effect of Customization Option Order on Usage . . . 123
A.13 News Customization at Different Times of the Day During the Post-Incentive Period . . . 126
A.14 News Customization by Experimentation . . . 128
A.15 Comparison to Other Experimental Estimates of Effects . . . 136
B.1 Alignment of Core Facts . . . 143
B.2 Drivers of Spin Including Phrases . . . 148
B.3 Summary Statistics . . . 150
B.4 Balance Table . . . 151
B.5 Attrition . . . 152
B.6 Baseline Predictors of Reported Willingness to Download the App . . . 153
B.7 Incentive-Period Exposure to Focal News Categories . . . 158
B.8 Effects on News Consumption . . . 166
B.9 Effects on Perceptions and News Satisfaction . . . 167

B.10 Midline Perceptions and Later News Consumption . . . 168
B.11 Effects on Knowledge and Polarization . . . 169
B.12 Downstream Outcomes and Multiple-Testing Adjustments . . . 169
B.13 Issue-Polarization Components . . . 170
B.14 Open-Ended Response Coding Scheme . . . 179
C.1 SCE sample construction . . . 183
C.2 Benchmark expectation inputs in the exact-12 panel . . . 183
C.3 SCE survey instruments and measures: benchmark and spending . . . 185
C.4 SCE survey instruments and measures: secondary and diagnostic variables . . . 186
C.5 ECB CES survey instruments and measures . . . 187
C.6 Illustrative mechanisms behind joint expectation movements . . . 188
C.7 Validation hierarchy and interpretation boundaries . . . 191
C.8 Focal-response and finite-correlation diagnostics . . . 192
C.9 PCA loadings for the six within-household correlations . . . 194
C.10 SCE reduced measures across sample and transformation variants . . . 194
C.11 SCE inflation-elicitation robustness . . . 195
C.12 Inflation measurement noise and disattenuated inflation links . . . 196
C.13 Primitive-correlation reducibility to focal response controls . . . 197
C.14 Auxiliary probability diagnostics for the primary SCE measure . . . 197
C.15 SCE sample-window robustness for signed-comovement means . . . 198
C.16 SCE panel-length audit . . . 198
C.17 Retained and excluded SCE respondents . . . 199
C.18 Permutation null benchmark for mean correlations . . . 199
C.19 SCE timing-window diagnostic . . . 199
C.20 Reliability of reduced-dimension comovement measures . . . 200
C.21 True-score rank calibration for the probability signed-comovement index . . . 200
C.22 Within-system probability revision diagnostic . . . 201
C.23 Descriptive reducibility to observed controls . . . 201
C.24 Descriptive correlates of probability signed comovement . . . 202
C.25 Inflation-density uncertainty and probability signed comovement . . . 204
C.26 ECB CES primitive adverse-comovement correlations . . . 205
C.27 ECB CES inflation-elicitation robustness . . . 206
C.28 ECB CES adverse macro-comovement by country . . . 206
C.29 ECB CES timing and country explanatory power . . . 207
C.30 Exploratory SCE clustering diagnostics . . . 207
C.31 Spending analysis sample . . . 208
C.32 Reduced-form spending relevance of the primary SCE comovement measure . . . 209
C.33 Secondary spending regressors . . . 209
C.34 Panel robustness for spending . . . 210
C.35 First-half to second-half spending sample . . . 210
C.36 First-half comovement and future spending . . . 211
C.37 Spending-slope interaction robustness . . . 212

C.38 Household bootstrap for the spending-slope interaction . . . . . . . . . . . . . . . . 213
C.39 Personal-risk prediction checks . . . . . . . . . . . . . . . . . . . . . . . . . . . . 214

# Introduction

Digital technologies and artificial intelligence are changing how economic actors produce, distribute, and consume information. They lower the cost of generating content and of tailoring how information is presented to individual preferences. At the same time, digital data make it possible to observe, at high frequency, how people interact with information and how their beliefs are organized over time. These changes create new opportunities to reduce frictions in information environments, but they also raise new questions about incentives, trust, the structure of economic beliefs, and the design of digital information environments. This dissertation studies these questions in three chapters on digital economics and AI.

The chapters are connected by a common question: how do information environments shape what people choose to read, what they trust, and how they organize economic beliefs? The first two chapters study the production and presentation of news, using a custom-built AI news app to generate controlled versions of the same stories. Chapter 1 studies customization, where users actively choose how stories are presented. Chapter 2 studies political spin, where random assignment to fixed article versions isolates framing while holding facts and coverage fixed. The third chapter turns to reception, using panel data on the same households over time to study how macroeconomic beliefs are connected within individuals. Across the three chapters, the dissertation separates parts of the information environment that are usually observed together: presentation from coverage, framing from facts, and the structure of expectations from their levels. A common methodological thread runs through all three, as each constructs a validated measure—of news preferences, of political spin, or of the structure of macroeconomic beliefs—that is not directly reported in standard data.

**Chapter 1: News Customization with AI** The first chapter studies whether AI can improve the match between consumers' news preferences and the news they consume. The chapter introduces a custom-built AI news app that allows users to customize the presentation of news articles while holding the underlying news event fixed. This design separates preferences over coverage from preferences over presentation, such as complexity, sentiment, opinion content, entertainment, and political perspective. The app generates these versions in real time from articles by a highly credible source, the Associated Press, and a fact-level alignment procedure together with a human validation survey confirms that the rewrites hold the underlying event fixed while meaningfully shifting the targeted dimension. A pre-registered experiment first shows that customization is valued at the adoption stage, as offering the feature significantly increases the rate at which people download the app. In large-scale field experiments, the chapter then documents substantial demand for customization and shows that users actively select article versions that better match their stated preferences. A central finding is that many users prefer less opinionated and more fact-driven news, and that customization increases news satisfaction without detectable effects on factual knowledge or polarization. Without customization, by contrast, users consume news that is far less aligned with their stated preferences, revealing a wedge between what people choose when article attributes are transparent and what they consume when they are not. This gap is starkest for negative and politically slanted content: users offered customization consume almost none of it, whereas those assigned a fixed version at random consume far more and report lower satisfaction, suggesting that choices made without customization can be a poor proxy for welfare. The chapter contributes to the economics of media markets by showing how AI can expand the news product space and reduce matching frictions between consumers and information providers.

**Chapter 2: The Asymmetric Response to Political Spin** The second chapter studies how political spin affects news consumption and political attitudes. Using the same AI news app, the chapter randomly assigns respondents to receive liberal, conservative, or politically neutral versions of the same news stories. Because the underlying news events are held fixed, the design isolates the causal effect of political framing from differences in story selection or outlet identity. Natural-language-processing tools confirm that the rewrites shift the political slant of articles as intended while leaving the underlying facts unchanged. Respondents are not told that different political versions of the same stories exist, so any updating about slant or credibility must come from the reading experience itself rather than from labels or beliefs about outlet identity. This design provides a direct test of demand-side theories of media bias, which predict that consumers reward aligned framing with greater engagement and satisfaction while penalizing misaligned framing. The results show an asymmetric response to political spin: respondents reduce consumption and downgrade credibility when exposed to politically misaligned spin, but they do not increase consumption or satisfaction when exposed to politically aligned spin. This asymmetry is concentrated among ideologically committed respondents, who disengage from misaligned spin rather than gravitating toward aligned spin. Downstream effects on learning and polarization are limited, though aligned spin modestly increases issue polarization. The chapter contributes to the literature on media bias by distinguishing the effects of framing from the effects of coverage and by showing that consumers punish misaligned spin more than they reward aligned spin. Because aligned spin does not draw consumers in, the chapter suggests that the persistence of political spin is more likely driven by supply-side incentives than by consumer demand.

**Chapter 3: Measuring the Structure of Household Macroeconomic Expectations** The third chapter shifts from the production and presentation of information to the observable structure of reported beliefs through which households interpret economic information. Using panel data from the New York Fed Survey of Consumer Expectations, the chapter studies whether households' macroeconomic expectations move independently or exhibit stable within-household comovement. It starts from within-household comovement among four expectations: inflation, unemployment, interest rates, and stock prices. The validated main measure is a signed-comovement index that ranks households by how tightly unemployment, saving-rate, and stock-price probability expectations move together over time, while inflation is the least connected component of the expectation system. The results show substantial heterogeneity in this expectation structure: for some households, macroeconomic expectations move as a connected system, while for others they behave more independently. Comparable comovement appears in the ECB Consumer Expectations Survey under different questions, response scales, and macroeconomic conditions. The measure is well separated from statistical noise, stable enough to rank households, and only partly explained by expectation levels, demographics, or response style. This structure matters for interpretation, as the association between inflation expectations and expected spending is flatter among households with more tightly connected probability expectations. For central banks and researchers, this implies that the same reported inflation expectation can carry different behavioral content depending on how it comoves with the rest of a household's macroeconomic expectations. The chapter contributes to digital economics by showing how panel data can reveal latent structure in economic beliefs that is not visible from expectation levels alone.

The dissertation's broader argument is that digital technologies and AI matter because they reshape the environments in which information is produced, presented, trusted, and interpreted. In news markets, AI expands the feasible product space and allows consumers to choose presentation styles that better match their preferences. In political information environments, the same technology makes it possible to isolate the effects of framing while holding facts and story selection fixed. In household expectations, digital panel data reveal that economic beliefs differ not only in their levels but also in their structure. Across these settings, the dissertation shows that the effects of information technologies depend not only on technical capabilities, but also on incentives, user choice, and the ways people organize economic beliefs.

**Author Contributions** Chapters 1 and 2 are joint work with Felix Chopra, Ingar Haaland, Christopher Roth, and Vanessa Sticher. All authors contributed equally to the conception of the projects, the experimental design, the interpretation of the results, and the writing of the manuscripts. My specific contributions included programming and maintaining the custom-built AI app, implementing the technical infrastructure for the experiments, and developing the fact-consistency validation exercise used to evaluate the AI-generated news articles. In Chapter 2, I additionally contributed to the NLP-based analysis of political spin and conducted additional empirical analyses. Chapter 3 is single-authored by me.

Chapter 1

# News Customization with AI

*joint with Felix Chopra, Ingar Haaland, Christopher Roth, and Vanessa Sticher*

**Abstract:** We introduce a new method to study news preferences by unbundling presentation from coverage. In our AI-powered news app, users can customize article characteristics, such as complexity or extent of opinion, while holding the underlying news event constant. In large-scale field experiments, we find that customization leads to better matching between the news consumed and stated preferences, increasing news satisfaction. While a significant fraction of users demand politically aligned news, the majority of users display high and persistent demand for less opinionated and more fact-driven news. By contrast, users not offered customization keep consuming news misaligned with their stated preferences.

*Keywords:* News Consumption, Customization, Artificial Intelligence, Matching

*JEL codes:* C93, D83, L82, P00

## 1.1 Introduction

A majority of Americans report dissatisfaction with the news media (Gottfried, 2020), a trend linked to declining news engagement (Forman-Katz, 2023) and broader negative social and political consequences (Prior, 2007; Sunstein, 2018; Zhuravskaya et al., 2020). This dissatisfaction is often attributed to how news is presented, including its negativity, political slant, technicality, and opinionated tone, suggesting a substantial mismatch between consumer preferences and the current supply of news.

Although markets often align supply with consumer preferences, there are several reasons to expect a persistent mismatch in news markets. News articles are multidimensional experience goods: fit is only revealed after consumption, and key attributes such as factual accuracy are costly to verify even ex-post. Because consumers make frequent, low-stakes choices over many potential articles, identifying content that matches preferences is very difficult (Gentzkow, 2018), making matching frictions especially pronounced in online news markets. On the supply side, increasing reliance on advertising revenue and referrals from social media networks create incentives to optimize for short-term engagement rather than consumer satisfaction, leading to clickbait content that attracts clicks but might cause regret (Kleinberg et al., 2024).

These market dynamics highlight the potential for a better alignment between the supply of news and consumer preferences. A precondition for improving the alignment is understanding consumers' news preferences: How do they actually want the news to be presented, and how much heterogeneity is there between consumers? To better understand these preferences, we need a setting where people actively choose characteristics of news articles, such as their level of opinion or complexity, while the underlying news events remain fixed.

To create such a news environment—where presentation is unbundled from coverage—we custom-built our own AI-powered news app, *TailorMadeTimes*. This app enables users to tailor news articles along *one* of the following dimensions at a time: extent of opinion, entertainment, sentiment, political perspective, and complexity.[1] The customization dimensions included in our app are directly motivated by the key sources of dissatisfaction about the prevailing news supply expressed by consumers (Reuters, 2023). As users enter their preferences with a set of sliders, the app dynamically presents AI-created rewrites, customized in real time, while holding constant the news event based on articles from a highly credible source, the Associated Press (see Figure 1.1 for an illustration of the app).[2]

By fixing the underlying news event and allowing users to vary one dimension at a time, the app cleanly separates presentation from coverage, an option largely absent in today's news market.[3] This unbundling may lower search costs and expand the product space, enabling combinations such as fact-driven reporting on highly politicized topics or non-technical treatments of economic news. The active choice environment, where all article attributes are transparently labeled, ensures that consumers know in advance what they will receive, reducing

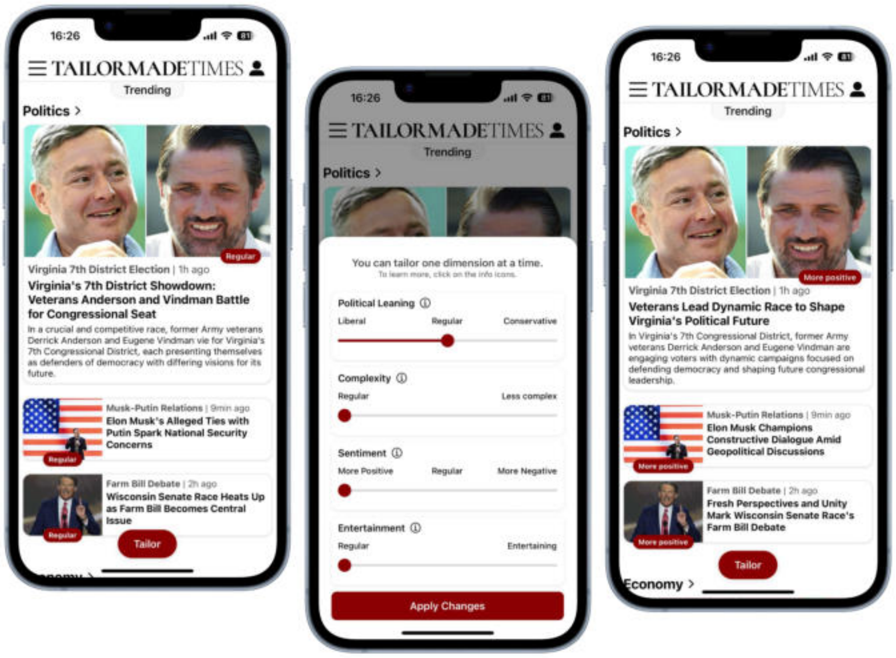
[1]To verify that the underlying news event is held constant across article rewrites, we develop a custom fact-level alignment procedure. Furthermore, we use a human validation survey to confirm that the AI rewrites meaningfully vary along the intended dimensions.

[2]In particular, the core factual elements (actors, outcomes, statistics, etc.) stay identical across versions.

[3]Naturally, news customization at scale has only very recently become technically feasible through recent advances in Generative AI.

the likelihood of ex post regret.

Figure 1.1: The AI News App




*Note:* This figure presents a mockup of the AI news app. The left screenshot shows the home screen, which comprises the top three trending articles per section ('Politics', 'Economy', 'US', 'World', 'Lifestyle', 'Sports') and is set to the *Regular* version. The middle screenshot shows the menu that opens when clicking the *Tailor* button, where users can customize articles via sliders. The right screen is the same as the first, but shows the articles in the *More Positive* version. The screenshots were taken on October 25, 2024, from an account for which we enabled all customizations except for *Facts Only*.

Leveraging our AI-powered news app in a large-scale field experiment, we examine the following main research questions: How much heterogeneity is there in news preferences across consumers? And does news customization help consumers improve the alignment between their news consumption and their stated preferences?

**Latent demand for customization** Before turning to our main experiment, we first conduct a pre-registered experiment with 5,005 U.S. respondents recruited from Prolific to test whether there is latent demand for language customization.[4] Participants were randomly assigned to learn about an AI news app either with or without the customization feature. Offering customization increases download rates by around 40 percent ($p < 0.01$), suggesting that users value the ability to tailor language features. We find similar download rates irrespective of whether the additional customization feature allows for political rewrites or the removal of opinions.

**Measuring news preferences** Having uncovered latent demand for customization, we next characterize heterogeneity in people's news preferences and study whether people use

[4]While there is a concern about LLM-powered bots in online surveys, we take several steps to mitigate this issue in our experiments. Most importantly, our main outcome—verifiable app downloads—requires a device with access to either the App Store or Google Play and is thus robust to even highly sophisticated bots.

customization to align their news consumption with their preferences. This requires a different design with the following features: Experience with the customization feature, a naturalistic setting, holding the set of news events covered in the app constant, and comparable engagement over time across treatment arms with and without customization. We implement such a design in our main experiment.

By holding the underlying news events constant and allowing deliberate variation in article features—such as complexity, extent of opinion, or political slant—we isolate specific dimensions of news preferences. At the same time, our approach of measuring preferences using real consumption data in an app directly captures the real world behavior of interest and mitigates concerns about social desirability bias, which confound self-reported data on news preferences.

Our experiment proceeds across multiple survey waves with U.S. respondents recruited from Prolific. In the initial recruitment survey, respondents are asked whether they are interested in downloading an AI-powered news app. Those with an interest are later invited to an app download survey in which they can download our app. Participants then use the app for 10 days with usage incentives, followed by midline and endline surveys measuring downstream outcomes.

We employ a between-subjects design, randomly assigning respondents to one of two customization treatments or to a control group without customization. Respondents only learn about customization after app login, ensuring no differential selection into app download by treatment status. In both customization groups, respondents can customize news articles along four dimensions. Three of the four dimensions are identical in both customization treatments, while the fourth dimension differs across treatments. Specifically, respondents in both customization treatments can choose between a regular default version or customize the articles in terms of entertainment (*Entertaining*), sentiment (*More Positive*, *More Negative*), and complexity (*Less Complex*), while holding the underlying news event constant. For the fourth dimension, we cross-randomize whether respondents can customize the political perspective of the articles (*Liberal*, *Conservative*) or remove opinion content (*Facts Only*).

To isolate the effect of being assigned different article versions without being offered customization, respondents in the control group are randomized into receiving only one of the eight article versions without customization. Articles are not labeled on the app for control group respondents, who are thus unaware of which version they receive.[5]

**How people use customization** A precondition for successfully characterizing people's news preferences using this experiment is that respondents actually use the news app. We therefore start by verifying compliance with the usage incentives. Respondents across all different treatment conditions spend a similar amount of time on the app. On average, they spend about four minutes per day and use the app on between 6 and 7 days during the 10-day incentive period. Taken together, our usage data confirm that people complied with the incentives. This ensures that users have experience with the customization tool and news articles from different rewrite versions.

Having established that usage incentives successfully encouraged respondents to use the app,

[5] Control group respondents are also unaware that other respondents can customize the news.

we now turn to our first main research question, drawing on data from the incentive period:[6] How much heterogeneity is there in news preferences across consumers? In the customization treatment with political rewrites, the most frequently used option is the politically more neutral *Regular* rewrite (37% of the time), but substantial time shares are also devoted to politically aligned news: Democrats spend 24% of their time in the *Liberal* setting and Republicans spend 25% in the *Conservative* setting. The *More Negative* option is rarely selected (2%), consistent with users making choices in the moment that are inconsistent with what they actually want (Kleinberg et al., 2024). These patterns indicate that when political rewrites are available, a significant fraction of our respondents use customization to align news with their ideological views. Our evidence is thus consistent with prior work showing that a subset of consumers have a preference for politically aligned news (Gentzkow and Shapiro, 2010). However, an important difference in our setting is that users make a deliberate choice to align the news with their own ideological views. Furthermore, while we find that a subset of consumers demand slanted news, we see that the large majority, in fact, prefer news without politically slanted content.

To assess whether the observed demand for the *Regular* setting reflects a genuine preference for politically balanced reporting or a default effect, we turn to the second customization treatment, which replaces the political rewrite options with a *Facts Only* option explicitly framed as "less opinionated and focused on the facts." In this treatment, which keeps the *Regular* version as the default, respondents spend 46% of their time in *Facts Only*—more than in any other setting—with only 18% of the time devoted to the *Regular* default. The pattern persists across partisan lines, is particularly pronounced among those with low baseline news satisfaction, and is equally pronounced among respondents with high and low baseline news consumption. This evidence suggests that stated preferences for more fact-driven and less opinionated content reflect people's true preferences, and not just social desirability bias.

**Customization and alignment with stated preferences** Having established significant heterogeneity in news preferences across consumers, we next examine our second main research question: To what extent news customization improves alignment between people's news consumption and stated preferences? We leverage pre-treatment questions from the recruitment survey in which respondents reported whether they find the news too complex, negative, positive, liberal, conservative, opinionated, or insufficiently entertaining. We then relate these assessments to subsequent consumption patterns during both the incentive and post-incentive periods. The results reveal that individuals systematically select news that corrects the perceived mismatch between existing news content and their preferences. For instance, respondents who think the news is too opinionated are significantly more likely to consume the *Facts Only* version whereas those who think the news is too complex spend more time using the *Less Complex* setting.

To examine whether the improved alignment between news consumption and preferences is unique to the customization treatments, we next study similar correlations between the perceived mismatch and news consumption among our control group respondents who are randomly assigned a fixed version throughout the whole period. For this exercise, we focus on the post-incentive period when control group respondents can freely adjust their consumption.

[6] As we discuss later, we obtain very similar results in the post-incentive period.

While control respondents who believe the news is too opinionated spend significantly more time using the app if they are assigned the *Facts Only* version, we do not observe significant correlations for any of the other dimensions. These findings demonstrate that customization strongly helps increase alignment with news preferences compared to being assigned a fixed version at random. Furthermore, the notable exception for the *Facts Only* version reinforces our earlier finding that there is a strong demand for more fact-driven and less opinionated news.

We next examine whether there are systematic group differences between treatment and control group respondents in their overall news consumption during the post-incentive period. While all respondents reduce their consumption in the post-incentive period, there are no treatment differences in the time spent using the app. Any treatment differences in news consumption will thus be driven by different time shares across article versions. The results reveal a striking *customization wedge* in news consumption: Respondents offered customization overwhelmingly consume fact-driven and less opinionated news and consume virtually no negative rewrites. By contrast, control group respondents spend very similar amounts of time across the different fixed article versions, making them much more likely to consume negative news in particular.

**Customization and news satisfaction** This customization wedge raises the question of whether a higher alignment with news preferences among respondents offered customization also raises news satisfaction. Consistent with an increased preference alignment, we observe significantly higher news satisfaction among respondents offered customization. This effect is particularly pronounced among respondents offered customization with the *Facts Only* option, consistent with the high demand for more fact-driven content. Looking into potential mechanisms, we find that the lower news satisfaction among control group respondents is largely driven by respondents assigned to more negative or politically slanted news. The fact that control group respondents consume the different article versions in roughly similar amounts, yet report lower satisfaction, implies that observed choices absent customization may be a poor proxy for welfare—consistent with the theory in Kleinberg et al. (2024) and echoing empirical evidence from social media on the consumption of toxic content (Beknazar-Yuzbashev et al., 2025).

Although customization improves alignment with preferences and positively affects news satisfaction, we find precisely estimated null effects on political polarization and factual knowledge. These null results are in line with a broader literature documenting muted downstream effects from news interventions, including those that alter ideological alignment or total consumption. Given that we hold the coverage of news events constant, it is perhaps unsurprising that our intervention does not meaningfully affect political polarization or factual knowledge.

**Related literature** We contribute to a growing empirical literature on people's news preferences (Braghieri et al., 2026; Chopra et al., 2022, 2024; Thaler, 2024). Robertson et al. (2023) show that negative sentiment drives online news engagement using large-scale field experiments, and Beknazar-Yuzbashev et al. (2025) find that removing toxic content from social media feeds reduces overall consumption. Our main contribution is to introduce a new approach for studying news preferences that unbundles content from coverage. Leveraging this approach, we show that most consumers—when offered an active choice over transparent product attributes—prefer fact-driven over opinionated news. These preferences are challenging to identify from market data because news articles are experience goods and outlets maximize engagement rather than user satisfaction (Kleinberg et al., 2024). Our finding that consumers show a large preference for *Facts Only* rewrites is consistent with models emphasizing the role of accuracy concerns when there is uncertainty about news quality (Gentzkow and Shapiro, 2006). Moreover, the significant wedge we document between news preferences under customization and those observed without customization aligns with recent evidence from social media suggesting that making more deliberate choices may substantially increase the reliability and reduce the slant of consumed news (Braghieri et al., 2026). Our evidence on the substantial demand for customization with *Facts Only* rewrites contrasts with the well-documented popularity of opinion shows in the US. This apparent discrepancy is consistent with recent evidence suggesting that consumers perceive opinion-oriented content as highly informative (Bursztyn et al., 2023), especially when it aligns with their own political views (Mitchell et al., 2018a).

Our study contributes to a large literature on the economics of media markets (Demsetz and Lehn, 1985; Mullainathan and Shleifer, 2005), media bias (Caprini, 2023; DellaVigna and Kaplan, 2007; Gentzkow and Shapiro, 2010; Groseclose and Milyo, 2005), and recent work that studies how artificial intelligence technologies and algorithms impact the market for news (Acemoglu, 2021; Acemoglu et al., 2024; Aridor et al., 2024, 2025; Cheng et al., 2025). Previous research has highlighted that AI technologies may increase the spread of misinformation (Acemoglu et al., 2026; Campante et al., 2025; Chen and Shu, 2024; Zhou et al., 2023). We highlight how AI technology can be used to address a market failure by increasing the match between consumer preferences and the supply of news.[7]

More broadly, our paper contributes to an emerging literature on how platform design choices shape digital news markets. Prior studies have focused on content moderation policies (Jiménez-Durán, 2023; Jiménez-Durán et al., 2024) and recommendation algorithms on social media platforms (Acemoglu et al., 2026; Beknazar-Yuzbashev et al., 2024; Kalra, 2025; Levy, 2021). Levy (2021) shows that Facebook's algorithm systematically limits exposure to counter-attitudinal news, creating mismatches between users' potential interests and the content they are actually shown, and thereby reinforcing polarization. By contrast, we highlight how deliberate news customization in a news app affects consumption patterns in a very different context—steering readers toward more fact-driven and less opinionated content.

## 1.2 The AI News App

### 1.2.1 App

For the purposes of this paper, we developed a mobile news app called *TailorMadeTimes* (see Figure 1.1 for a mockup). The design and structure of our news app are informed by common best practices of other mobile news apps. Specifically, news articles are previewed in the format of a scrollable *home feed*, which is organized by news categories ("Politics", "Economy", "US",

[7] Methodologically, we relate to a growing literature that uses LLMs, e.g., in the context of the collection of qualitative data (Chopra and Haaland, 2026) or the classification of text (Ash and Hansen, 2023; Braghieri et al., 2024; Graeber et al., 2024; Habibi et al., 2026; Noy and Rao, 2025).

"World", "Lifestyle", "Sports"; in this order). Article previews include the article title, timestamp (e.g., "1 hour ago"), a short AI-generated topic label (e.g., "US Presidential Election"), and an image for most articles (during the study period, 82% of articles had an image). For the leading article in each section, we also display the article lead already in the feed. Every news category in the feed includes the top three trending articles, and users can switch to dedicated category sections with additional articles. The app is updated with new articles every hour, resulting in an average of 112 new articles per day, where articles remain in the app for three days. Older articles can still be accessed via the search and bookmark features. Furthermore, users with push notifications enabled receive four notifications throughout the day, each containing the most trending news story since the previous notification (see Section 1.2.2 for a definition of "trending").

The distinguishing feature from other news apps is that users of our app can *customize* news articles in terms of entertainment, sentiment, political perspective, complexity, and the extent of opinions—while holding the underlying news event constant. As shown in Figure 1.1, users can indicate their preferences by clicking the "Tailor" button, which activates a pop-up menu with a set of sliders. As users enter their preferences by changing the slider settings, the app dynamically updates *all* articles. Users can make changes to the sliders at any point in time. Importantly, the set of news articles remains identical: customization only affects the presentation of news articles along one dimension at a time, not which articles are shown in the app.

### 1.2.2 Source of Articles

The news app relies on news articles sourced from the Associated Press (AP), which is widely regarded as a highly credible news organization.[8] Established in 1846, AP has a long-standing reputation for its strict adherence to journalistic ethics and its commitment to unbiased news coverage. It operates on a not-for-profit model, which allows it to focus solely on news gathering and reporting. As a result, its news is trusted by a vast array of subscribers and outlets around the world. For our app, we select English-language news articles covering politics, economy, the US, the world, lifestyle, and sports via a news API subscription (www.newsapi.ai), focusing on articles that have between 100 and 1,000 words. Then, for each article, we create a *trending score* to capture its popularity. Specifically, we retrieve a score from the news API that indicates how often entities within an article (e.g., people, events, countries) have been mentioned across all available news sources within the past two days before the article was published. We combine this score with a measure of recency and sort articles by this combined score, such that articles that contain trending entities and that are more recent appear higher up within the home and category feeds of the app.

[8]The Associated Press is consistently rated as one of the news agencies with the highest reliability scores (Bhadani et al., 2022) and is also among the most trusted news sources in the US (see, e.g., https://today.yougov.com/politics/articles/52272-trust-in-media-2025-which-news-sources-americans-use-and-trust.

### 1.2.3 Article Versions

This section motivates and characterizes the different article versions we offer users as part of the customization feature of the app (via the "Tailor" button). The baseline version is the *Regular* version, which is a summary of around 300 words of the original AP article. The customization options on top of *Regular* directly correspond to common news reader complaints emphasizing the dryness, negativity, political bias, and technicality of mainstream reporting (Reuters, 2023). The rewrite dimensions we focus on (extent of opinion, political perspective, entertainment, sentiment, and complexity) directly reflect these key sources of dissatisfaction. Political and sentiment rewrites address concerns about political bias and negativity. Entertainment and complexity rewrites respond to complaints about dryness and technicality. *Facts Only* rewrites reduce opinionated content, offering an option for more neutral reporting. All rewrites vary the presentation while covering the same core news event (e.g., events, actors, dates, outcomes, statistics). Table 1.1 provides an overview of the rewrite dimensions with examples, illustrating how the same news event is expressed differently across these dimensions. Below, we describe each rewrite in more detail, explicitly referring to and directly quoting the examples presented in the table (the technical implementation of the rewriting process is described in Section 1.2.4).

**Political** Political rewrites adjust wording and framing to reflect liberal or conservative perspectives. For example, as shown in Table 1.1, the *Conservative* version frames the reforms positively by highlighting lawmakers expressing "thoughtful concerns" about President Trump's "ambitious" government reforms, led by the "prominent entrepreneur Elon Musk."[9] Conversely, the *Liberal* rewrite describes the same initiative skeptically, noting criticism for its potential overreach, reflected in its title emphasizing "Bipartisan Criticism." As another example, inflation data under a *Conservative* rewrite might be framed as a concern, such as "Inflation still above its target under the current Biden administration." By contrast, *Liberal* rewrites highlight progress and successes, such as "Inflation significantly down from its peak under the current Biden administration." More generally, political rewrites vary the use of partisan terminology, as in Gentzkow and Shapiro (2010), where, for example, the *Liberal* rewrite would refer to "estate tax," whereas the *Conservative* rewrite would say "death tax." These rewrites aim to reflect differing political perspectives without introducing unsupported causal claims.

**Sentiment** Sentiment rewrites shift the article's tone toward optimism or caution. As shown in Table 1.1, the positive rewrite emphasizes lawmakers engaging in "constructive discussions," framing the reforms as "ambitious," with the title highlighting "Collaboration" and "Greater Efficiency." In contrast, the negative rewrite describes lawmakers as "increasingly alarmed," emphasizing "chaos" and "backlash," underscored in the title as well.

**Entertainment** *Entertaining* rewrites employ dynamic and engaging language, using vivid descriptions, creative analogies, and storytelling techniques. For instance, Table 1.1 presents an engaging narrative by characterizing the situation as sparking GOP drama and describing Elon Musk's leadership as a "plot twist worthy of Silicon Valley." The title itself emphasizes drama to enhance reader interest. As another illustration, a straightforward phrase like "Persistent inflation" might be transformed into "Inflation refuses to leave the room," framing economic

[9] When this article appeared, Elon Musk worked for the government and was looked upon favorably by Trump.

Table 1.1: Rewrite Examples

| Dimension | Title and first sentence of body | Explanation |
|---|---|---|
| Regular | *Title*: Republican Lawmakers Push Back on Key Aspects of Trump Administration's Efficiency Reforms<br>*Body*: Republican lawmakers are voicing concerns about President Trump's sweeping government reforms under the Department of Government Efficiency (DOGE), led by billionaire Elon Musk. [...] | (Baseline Associated Press Summary) |
| Conservative | *Title*: Republicans Balance Support for Trump Reforms with Local Concerns<br>*Body*: Republican lawmakers are expressing thoughtful concerns about President Trump's ambitious government reforms spearheaded by the Department of Government Efficiency (DOGE), under the leadership of prominent entrepreneur Elon Musk. [...] | Downplaying issues, positive framing of Republican government plans |
| Liberal | *Title*: Trump's Government Reform Plan Faces Bipartisan Criticism for Overreach<br>*Body*: Republican lawmakers are voicing concerns about President Trump's sweeping government reforms under the Department of Government Efficiency (DOGE), an initiative spearheaded by billionaire Elon Musk that has faced significant criticism for its potential overreach. [...] | Highlights criticism and concerns associated with Republican government plans |
| More Positive | *Title*: Republicans Collaborate on Ambitious Government Reforms for Greater Efficiency<br>*Body*: Republican lawmakers are engaging in constructive discussions about President Trump's ambitious government reforms under the Department of Government Efficiency (DOGE), spearheaded by visionary entrepreneur Elon Musk. [...] | Optimistic framing, positive attributes |
| More Negative | *Title*: Trump's Sweeping Reforms Spark Chaos and GOP Backlash<br>*Body*: Republican lawmakers are increasingly alarmed by President Trump's sweeping government reforms under the Department of Government Efficiency (DOGE), headed by billionaire Elon Musk. [...] | Cautionary and alarmist framing |
| Entertaining | *Title*: Musk-Led Government Overhaul Sparks GOP Drama Over Cuts<br>*Body*: Republican lawmakers are raising eyebrows at President Trump's ambitious government overhaul helmed by the Department of Government Efficiency (DOGE), which, in a plot twist worthy of Silicon Valley, is being led by billionaire entrepreneur Elon Musk. [...] | Uses dramatic and vivid language |
| Less Complex | *Title*: Republicans Worry Musk-Led Government Changes May Hurt Jobs and Programs<br>*Body*: Some Republican lawmakers are worried about President Trump's major changes to the government. These changes are being managed by the Department of Government Efficiency, or DOGE, which is led by billionaire Elon Musk. [...] | Simpler wording, shorter sentences, easier syntax |
| Facts Only | *Title*: Republican Lawmakers Raise Concerns Over DOGE Government Reforms<br>*Body*: Republican lawmakers have expressed concerns about President Trump's government reforms under the Department of Government Efficiency (DOGE), led by Elon Musk. [...] | Objective and neutral phrasing |

*Note:* This table provides an overview of the rewrite dimensions with examples of how article title and first sentence of the body vary across dimensions. The example shown is from an article published on February 12, 2025. Text highlighted in yellow indicates examples of dimension-specific rewrites. The full article body along with title and lead are available on the following link: https://raw.githubusercontent.com/cproth/papers/master/TMT_examplerewrites.pdf.

issues with a touch of humor or drama. Storytelling techniques may frame developments as dramatic arcs, such as likening market fluctuations to a "roller-coaster ride." Relatable metaphors or analogies, such as comparing inflation to "an uninvited guest overstaying its welcome," further enhance reader engagement.

**Complexity** *Complexity* rewrites simplify language and structure. These rewrites adopt simpler vocabulary and shorter sentences, similar to the style of the *Simple English Wikipedia*. Table 1.1 provides a direct example: "Some Republican lawmakers are worried about President Trump's major changes to the government. These changes are being managed by the Department of Government Efficiency, or DOGE, which is led by billionaire Elon Musk." This example demonstrates how a long sentence can be effectively split into two shorter sentences, each using simple sentence structures and accessible vocabulary such as “worried” and “changes” instead of more complex terms. As another example, the sentence “Signs of decreasing inflation and a cooling job market have led to expectations of monetary easing” might be rewritten as “Because inflation is going down and the job market is slowing, people think the central bank will lower rates.” Additionally, technical terms are either replaced with simpler terms or explained in simple language, such as defining “monetary easing” as “making it easier for businesses and people to borrow money.”

**Facts Only** *Facts Only* rewrites reduce opinionated content by removing subjective assessments. Table 1.1 exemplifies this approach with a neutral description: “Republican lawmakers have expressed concerns about President Trump’s government reforms under the Department of Government Efficiency (DOGE), led by Elon Musk.” The title also neutrally states that lawmakers “Raise Concerns” without subjective framing. As another example, the sentence “The government’s reckless spending has led to an unsustainable increase in national debt, putting the economy at risk” might be rewritten as “The government’s spending has increased the national debt.” These rewrites eliminate subjective descriptors (e.g., “reckless” or “unsustainable”), present the facts objectively, and simplify the statements. Such rewrites present the news objectively, fostering unbiased reporting by removing evaluative language.

### 1.2.4 Multi-Agent AI System for Rewrites

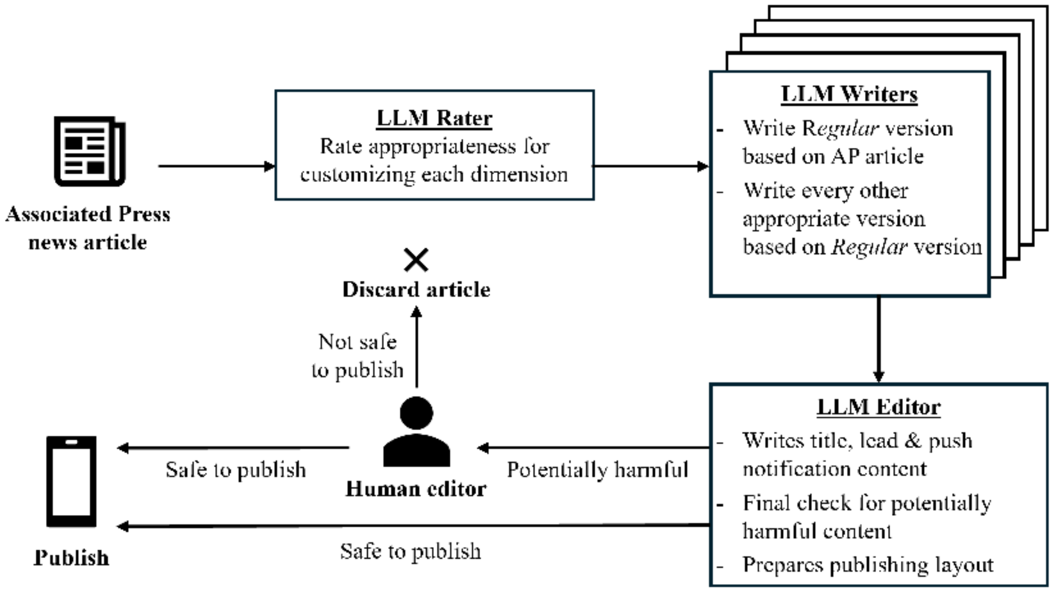
To systematically vary article characteristics, we built a multi-agent AI system. The objective of this system is to generate alternative versions of each AP news article that hold the underlying news event constant and uphold journalistic integrity across versions by not omitting or fabricating facts. To achieve this, we use a variety of AI agents for specific generative and evaluative tasks. Our system is built on the Generative Pre-trained Transformer 4o model family (“GPT-4o” and “GPT4o-mini”) by OpenAI, which was state-of-the-art at the time of the experiments. Figure 1.2 illustrates this workflow, which consists of three components outlined below.

**Preprocessing** Starting with a given AP news article (sourced as described in Section 1.2.2), the first agent assesses the appropriateness of customizing the article. Specifically, for every dimension we set out a rule for when it should not be customizable. For example, articles that contain sensitive topics, such as death or terror, cannot be made more entertaining or more positive. Further, it also flags articles categorized as “Sports” or “Lifestyle” to not be customizable on the political dimension, given the limited scope for political rewrites for this type of content.[10]

**Writing** Given the original AP article, an agent first writes the *Regular* version, which is a shorter, summarized version of the AP article of around 300 words. Based on this version, a first draft of all other appropriate versions is written. All rewrite agents are instructed to (i) maintain accuracy by holding the news event constant, (ii) maintain article length and logical flow, and (iii) not mislead, omit, or fabricate even in light of changing emphasis and framing.

[10] In the app, users receive an explanation via a tooltip on why a specific article-version combination is unavailable.

Figure 1.2: Multi-Agent AI System for Rewrites




*Note:* This figure illustrates the multi-agent AI system used to generate the rewrites of the articles. First, an agent assesses whether customizing the article is appropriate (preprocessing). Next, a writing agent generates the *Regular* version by summarizing the article and drafts the rewrites. An editor agent then creates titles, leads, and push notification content. Finally, the rewrites are screened through OpenAI's Moderation API and, if flagged, sent for manual review before publication.

Additionally, we equip every rewrite agent with relevant context beyond the model's cutoff date (e.g., who the current US president is, what today's date is, etc.). On top of these general guidelines, every agent is provided with a description of the desired characteristics of the rewrite and a version-specific curated set of input/output examples (*few-shot prompting*); the shared and version-specific prompts are documented in the shared prompts appendix (Sections E.1 to E.4, E.7 and E.9). By showing representative example rewrites and explanations of rewritten sentences, we effectively teach the agent how to implement a rewrite that corresponds to the desired characteristics. Lastly, we ask the agents to conduct the rewrite in a two-step process, where they first identify suitable sentences to be rewritten and provide a justification for each suggested sentence rewrite. Second, they are instructed to reflect on the suggested changes and implement those that are in line with the general rewrite guidelines and the desired dimension characteristics. This reflective procedure aims at maximizing output quality while adhering to the constraints we set.

**Editing** Lastly, an editor agent, instructed similarly to the rewrite agents for the main body, writes a version-specific title and lead. Further, it generates push notification content for articles that, based on the trending score, are chosen to be sent as a push notification to users. The push notifications correspond to the customization setting of the user when the notification is sent. As a final check, all content that has been generated by AI in the process is sent to OpenAI's Moderation API to rule out publishing unsafe content. If any content is flagged by the Moderation API, the entire article with all its versions is not published, but sent to us for review. In addition, all articles that received the lowest appropriateness score are also sent to us for review. Unflagged articles are automatically and immediately published on the app.

### 1.2.5 Validation of News Event Consistency

**Methodology** To validate that rewrites do not change the underlying news event, we developed a custom fact-level alignment procedure. We begin by extracting factual claims, defined as single, objectively verifiable statements following Ni et al. (2024), from every article version. Next, we classify these extracted facts into core and supplementary using an LLM classifier (GPT-4.1). Core facts are defined as those that (i) directly convey the article's main event, "who did what", and its primary significance ("why it matters"), and (ii) form the minimal sufficient set to reconstruct a faithful account of the core events. Supplementary facts provide context, background, or stylistic detail and are allowed to vary across rewrites (e.g., explanations added in the *Less Complex* version). We restrict validation to core facts because variation in supplementary details is expected by construction and does not change the underlying news event. This means checking whether all core facts from the *Regular* version are contained in the rewrites (to detect omissions) and vice versa (to detect hallucinations). To operationalize this, we align facts via embeddings: for each core fact in the *Regular* version (and vice versa), we take the maximum cosine similarity to any fact in the rewrite (core and supplementary) and declare a match if the maximum similarity $\geq \tau$, where $\tau$ is calibrated against human judgments on a held-out set ($\tau$=0.66). As a safeguard against missed matches, we then conduct a final LLM-based consistency check to reduce false negatives. In addition to the embeddings-based checks, we separately check the consistency of extracted numbers across versions (e.g., dates, statistics, percentages). A detailed account of the technical implementation of this three-step validation, as well as example outputs of all steps, is provided in Appendix A.2.

**Results** Table 1.2 reports directional coverage of core facts across rewrites across 59,092 article versions in total. We present two measures: $R \rightarrow V$ (share of *Regular* core facts covered by the rewrite; 100% minus the omission rate) and $V \rightarrow R$ (share of *Version* core facts supported by the *Regular* article; 100% minus the hallucination rate). The "Embeddings" block uses cosine similarity at the calibrated threshold of 0.66, while "+ LLM Adjudication" reevaluates the unmatched remainder of the embeddings-based analysis against the full target article to avoid potential false negatives. Lastly, "Numeric Alignment" assesses equality of extracted numbers (dates, quantities, percentages) across versions.

Across all comparisons, embedding-only alignment yields coverage of 82.4% (R→V) and 82.0% (V→R), indicating substantial baseline alignment. Under embeddings alone, coverage is lowest for *Entertaining* and *Less Complex* (≈ 72–78%), consistent with heavier paraphrasing and structural changes, and highest for political rewrites and *Facts Only* (≈ 86–90%). After LLM adjudication, coverage rises to 96.5% (R→V) and 96.3% (V→R). Under this adjudicated measure, our metrics are all above 90% in both directions, and five out of seven versions exceed 95% in both directions. Furthermore, quantitative content is preserved: numeric alignment is close to 100% in both directions for every rewrite, providing an orthogonal validation that dates, statistics, and percentages are consistent. Overall, the evidence indicates that rewrites retain the underlying news event: embedding similarity captures most matches; the LLM adjudication reliably recovers hard paraphrases; and numerical details remain consistent. For more details

Table 1.2: Alignment of Core Facts

| | Embeddings | | + LLM Adjudication | | Numeric Alignment | |
|---|---|---|---|---|---|---|
| *Version* | *R→V* | *V→R* | *R→V* | *V→R* | *R→V* | *V→R* |
| Conservative | 88.4% | 87.5% | 96.8% | 95.9% | 99.2% | 100% |
| Liberal | 90.0% | 87.7% | 98.0% | 95.8% | 99.6% | 100% |
| Entertaining | 72.0% | 73.5% | 97.5% | 97.7% | 98.3% | 99.7% |
| Less Complex | 76.7% | 78.3% | 98.2% | 99.4% | 98.8% | 99.0% |
| More Negative | 83.5% | 79.9% | 96.2% | 91.8% | 99.3% | 99.9% |
| More Positive | 82.3% | 83.2% | 91.3% | 95.3% | 97.7% | 100% |
| Facts Only | 86.4% | 87.4% | 96.2% | 95.8% | 99.9% | 100% |
| All | 82.4% | 82.0% | 96.5% | 96.3% | 98.9% | 99.7% |

*Note:* R=Regular, V=Version. "R→V" is the share of *Regular* core facts covered by the rewrite (1–omissions). "V→R" is the share of *Version* core facts covered by *Regular* (1–hallucinations). "Embeddings" uses `all-mpnet-base-v2` at the calibrated cosine similarity threshold ($\tau$=0.66). "+ LLM Adjudication" checks unmatched pairs with an LLM to rule out false negatives. "Numeric Alignment" reports consistency of extracted numeric tokens across versions. Examples of this alignment procedure can be found in Appendix A.2.

and examples, see Appendix A.2.

### 1.2.6 Human Validation of Rewrites

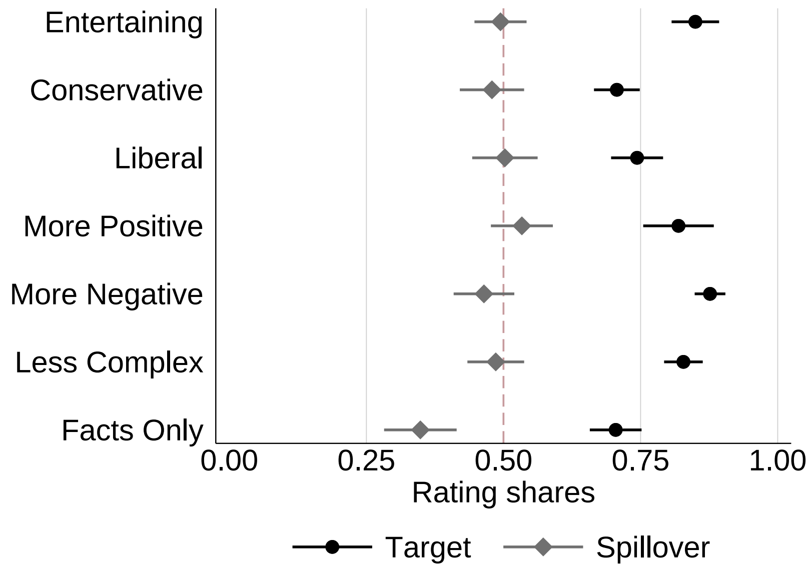
**Design** To validate whether and how the rewrites affected the articles, we recruited 1,169 respondents from Prolific (see Appendix Table A.1 for an overview of studies).[11] To ensure political balance in our sample, we stratified the collection to recruit as many Republicans as Democrats. The data collection based on articles from our wave 1 field experiment took place in December 2024, while the one for wave 2 took place in August 2025.[12] For each wave, we randomly selected 21 news articles among the trending articles from the politics section, which appear at the top of the app during the study period and are the most commonly read on the app. For these news articles, we then construct within-article version pairs, with each version pair consisting of the *Regular* version and a rewritten version.[13] Each respondent then rated six randomly selected version pairs along different dimensions. Version pairs are presented in randomized order. We provide respondents with the title, the article's lead, and the full article. To measure whether our rewrites affected the target dimensions as expected, we ask half of the respondents to rate the article pairs in the target dimension of interest. For example, if the target dimension is complexity, we ask our respondents "which article is more complex?" To assess whether our rewrites systematically induced spillovers to other dimensions, we assign the other half of respondents to rate the article pair along a randomly chosen non-target dimension (e.g., "which article is more positive?"). This design allows us to examine whether we effectively changed articles only along the target dimension or also along other dimensions.

[11] The experimental instructions are available here: https://raw.githubusercontent.com/cproth/papers/master/tmt_instructions_appendix.pdf.

[12] Given increasing concerns about bots on online survey platforms (Rilla et al., 2026), we took additional measures in the wave 2 survey. In particular, participation in our survey required authentication with a mobile phone, where each physical device could unlock the study only once.

[13] In wave 1 of the human validation, we did not include the *Facts Only* rewrite. To ensure overall similar sample size across the waves, we over-sampled *Facts Only* rewrites in wave 2 of the human validation by a factor of 2.

Figure 1.3: Human Validation of Article Rewrites




*Note:* This figure shows the share of respondents rating the rewritten version of the article as being more entertaining, conservative, liberal, positive, negative, or less complex compared to the regular version of the article in a blinded binary comparison. The black dots show the fraction of respondents correctly identifying the article version that was varied along the target dimension (e.g., *Entertaining* rewrites being rated as more entertaining). The gray squares show the fraction of respondents classifying an article as belonging to a spillover dimension (e.g., non-*Entertaining* rewrites being rated as more entertaining). Error bars represent 95% confidence intervals.

**Results** The black circles in Figure 1.3 separately present the fraction of respondents that correctly classified the article for each of the seven different rewrite dimensions. The figure highlights that the fraction of human raters who correctly classify the articles is quite high on average. Around 80% of raters correctly identify the *Entertaining*, *More Positive*, *More Negative*, and *Less Complex* rewrites. For the political rewrites (*Liberal* and *Conservative*) and *Facts Only*, the fraction of correct classifications is slightly lower at around 70%. We document similar human ratings of the article pairs for raters who identify as Democrats and Republicans (Appendix Figure A.1).

The gray squares in Figure 1.3 show the fraction of respondents who classified the article as belonging to a given spillover dimension. For example, the gray square for *Entertaining* means that articles rewritten in all dimensions other than *Entertaining* are, on average, classified as the *Entertaining* version in 49% of the time. If non-target dimensions were not systematically affected by rewrites, we would expect this number to be close to 50% if respondents select a version at random in the absence of a perceptible difference. We find similar patterns in all other spillover dimensions and cannot reject the null hypothesis that rewrites did not systematically affect individual non-target dimensions. We observe similar human classification patterns in wave 1 and wave 2 of the survey.

Appendix Figure A.2 shows the spillovers to other rewrite dimensions separately for each target dimension. This figure reveals overall muted spillovers for each of our target dimensions, except for the *Facts Only* rewrite. In particular, *Facts Only* articles are rated as less negative and less liberal, consistent with these rewrites being more fact-driven and less opinionated. As such, being rated as less political and less likely to be negative is a natural byproduct of more neutral and fact-driven reporting. Appendix Figure A.30 provides additional evidence that respondents update their beliefs about the article characteristics after using the news app for 10 days in all dimensions except for entertainment.

#### 1.2.7 Ethical Considerations

We took several measures to mitigate potential concerns about AI-generated news content.[14] First, trained research assistants vetted the factual accuracy of rewritten versions compared to the baseline AP article during our study period. Second, our multi-agent AI system incorporates multiple quality checks, including screening for harmful content. Third, our AI writing assistants are explicitly instructed to uphold rigorous standards of journalistic integrity and to neither omit nor fabricate information. Fourth, our rewrites for conservative and liberal perspectives do not introduce unsupported causal claims and do not affect the underlying facts mentioned in the rewritten articles. Fifth, our news app only relies on articles from a high-credibility source, the Associated Press. Sixth, we employ a conservative rewriting system where an AI agent first reviews and then flags articles that would be potentially inappropriate to customize along a specific dimension (e.g., stories surrounding death or terror cannot be made more entertaining or more positive).

### 1.3 The Demand for Customization

Before addressing our main research questions, we first examine whether there is demand for news customization, as this would suggest that there is a mismatch between their preferences and the supply of news, which they think news customization can help address.

Examining whether people value news customization requires a design that isolates this demand from people's demand for the underlying news articles, since customization can only be offered as a bundle with the articles to which it is applied. To address this, we offer participants access to different versions of the app in a between-subjects design: one version with the option of customization and an identical version without customization. We then infer the demand for customization from differential download rates of different app versions. This revealed-preference measure captures the first stage of product adoption: whether people are willing to incur the cost of downloading a new app in order to gain access to news customization.

#### 1.3.1 Experimental Design

**Sample** On December 13 and 14, 2024, we recruited 5,005 Android and iPhone users via Prolific, a survey platform commonly used in social science research (Peer et al., 2022). Appendix Table A.4 presents summary statistics of our sample and Appendix Table A.5 examines the integrity of randomization. Our sample matches the US population in terms of age and gender, but includes a higher share of individuals with at least some college education (86% in our

[14] We received IRB approval from MIT, the Frankfurt School of Finance & Management, the University of Cologne, and the NHH Norwegian School of Economics.

sample vs 63% in the general population), which is a common feature of online panels (Armantier et al., 2017). 42% identify as Democrats, 33% as Independents, and 25% as Republicans. 84% primarily consume news online, which is very close to the 86% of Americans who report consuming news online.[15]

**Design** Appendix Figure A.3 provides an overview of the design. Participants are randomly assigned in equal proportions to one of three conditions: the control condition or one of the two customization conditions. All participants are informed about our AI news app with AI-written articles based on content from the Associated Press, a high-quality news provider. The instructions emphasize the factual accuracy and human review of AI-written articles to mitigate potential participant concerns; the full demand-experiment instructions are reproduced in the shared appendix (Section D.2).

Participants in the customization conditions *additionally* learn that the news app includes the option of customizing news articles in terms of three target dimensions (complexity, sentiment, and entertainment) without changing the facts included in the news article. Specifically, participants receive the following instructions:

> The key feature of our news app is that you can tailor the news stories according to your preferences. Click the Tailor button and move the sliders to choose how you want the news to be reported. [. . . ] You can change slider settings at any point in time, and the app will update the writing of all articles in real-time. Irrespective of how you tailor the news, the underlying facts remain constant. Tailoring also does not affect the selection of articles – it only changes how they are presented.

We explain the effect of each customization dimension to participants and demonstrate how customization would work in practice using a short video.

To assess whether demand is sensitive to the range of adjustable features, we vary the fourth dimension offered in the treatment conditions. In one variant, participants can additionally select the political orientation of articles (conservative versus liberal). In the other, they may opt to reduce opinion content and emphasize factual reporting.

> Changes the political perspective of the article by using language and terms that are more commonly used by either conservatives or liberals. Also changes the focus of the article to highlight themes and issues that matter most to each political group.

The rewrite dimension that makes articles less opinionated (*Facts Only*) is described as follows:

> Makes the article less opinionated by focusing on facts only. Also avoids subjective assessments and judgments, and uses more neutral language.

We then ask all participants whether they are interested in downloading the app. For those who express interest, we provide detailed download instructions, including a link to the app and a QR code for scanning.[16] We explicitly communicate to participants that payment for

[15] https://www.pewresearch.org/journalism/fact-sheet/news-platform-fact-sheet

[16] Android users can download the app directly from the Google Play Store, while iPhone users need to first download *TestFlight*—an app for beta testing—before being able to download our app. Respondents take between three to four minutes on average to download and log into the app.

survey participation will not depend on downloading the app. Our main outcome of interest is a dummy variable for whether participants actually download and log into the app.

### 1.3.2 Results

Figure 1.4 and Appendix Table A.6 show that news customization increases the download rates and logins into the news app compared to an otherwise identical news app. In particular, Figure 1.4 shows that the download rate in the customization group is 9.6% compared to 6.8% in the control group without customization ($p < 0.01$). The magnitude of this effect is economically large, corresponding to a 40% increase in download rates compared to the control group.

The customization conditions with political rewrites and *Facts Only* rewrites allow us to see whether there is a higher demand for customization that adds a political perspective or removes opinions. Interestingly, we find similar download rates across the two customization conditions ($p = 0.747$): 9.4% download the app when the customization allows for political rewrites, compared to 9.7% when the customization allows for less opinionated rewrites (as shown in Figure 1.4). This finding suggests either heterogeneity in demand — where some users prefer political alignment and others prefer more fact-driven content — or that both customization options are viewed as equally attractive.

In addition to measuring latent demand for the app, we can also track how time spent in the app differs between different treatments.[17] While app downloads are endogenous to the treatments, it is nonetheless interesting to examine whether those in the customization treatments spend more time in the app. As shown in Appendix Figure A.4, on the day of the app download, users spend significantly more time in the app when offered the *Facts Only* customization compared to no customization ($p = 0.067$). They also spend somewhat more time when offered political rewrites, although this difference is not statistically significant ($p = 0.371$). While the overall usage is rather low (users across conditions spend on average 125 seconds in the app on the first day, conditional on downloading it, and usage quickly drops thereafter), these results suggest that customization in general and the *Facts Only* option in particular lead to higher initial news engagement.[18] Given our discussion above, our first main result follows.

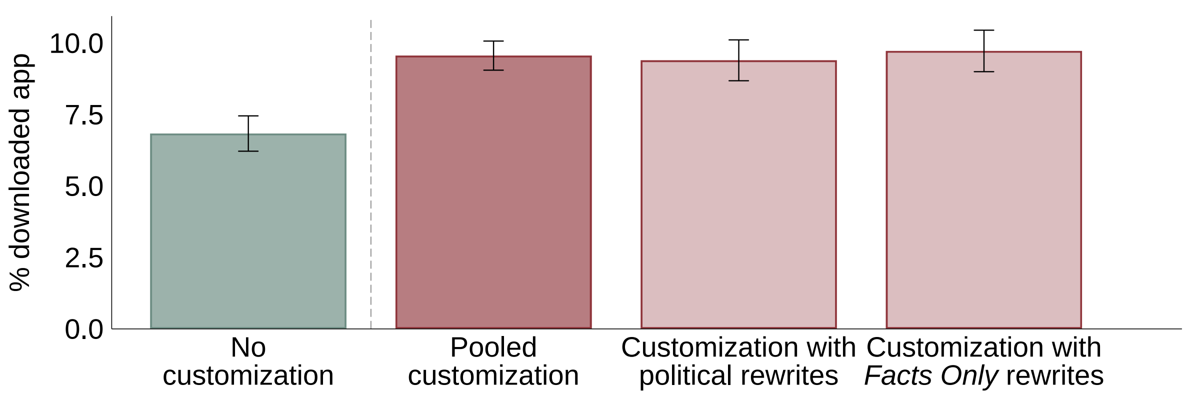
> ***Result.*** *We document a latent demand for news customization: respondents offered the app with customization are 40% more likely to download it than those offered the same app without the customization option. We find similar demand for customization with political rewrites and less opinionated rewrites, suggesting an underlying demand both for amplifying and minimizing opinion in the news.*

While this data underscores that there is latent demand for news customization, it only allows us to characterize user preferences for a small subset of the population and it does not allow us to study downstream outcomes with sufficient statistical power. Measuring news

[17]We do not directly observe if the app is open on a user's phone. Instead, we infer time spent from user actions (e.g., clicks, scrolls). Actions separated by less than ten minutes are grouped into a session and the session duration is defined as the time between the first and last action. Session durations are then aggregated (e.g., to the daily level).

[18]Appendix Figure A.5 presents average app usage frequency and time during the first seven days after download, confirming the results from day 1.

Figure 1.4: Demand for News Customization




*Note:* This figure shows the share of respondents who downloaded the AI news app and logged in at least once across conditions. The first bar shows the download rate in the no customization condition. The second bar shows pooled download rates for both customization conditions. The last two bars show download rates separately for each customization condition (with political rewrites and with *Facts Only* rewrites). Error bars represent standard errors of the mean.

consumption with customization in a broader population, and assessing the implications of such customization, requires an experimental design with two additional features: prior user experience with the customization tool and comparable levels of engagement across the conditions with and without customization. We turn to such a design in the next section.

## 1.4 News Consumption With Customization

Having established latent demand for news customization, we now turn to our main research question: measuring news preferences. In our main experiment, users are randomly assigned to customization or control, ensuring comparable exposure and enabling us to identify preferences.

### 1.4.1 Measuring News Preferences with Customization

Measuring consumers' news preferences is difficult for a variety of reasons. First, news articles vary across multiple dimensions, even when covering the same event. This multi-dimensionality complicates revealed-preference approaches that infer preferences from observed choices. If we conceptualize news articles as a vector of product attributes, then news consumption involves simultaneous trade-offs over multiple correlated attributes. If, for example, slanted news tends to be more vivid or sensational, then disentangling a preference for slant from a preference for a more vivid language becomes more challenging.

Second, news articles are *experience goods*—their quality and fit can only be assessed after consumption (Gentzkow, 2018). Since news outlets typically optimize for engagement rather than consumer satisfaction (Kleinberg et al., 2024), "clickbait" stories are supplied even though consumers might regret reading them ex post. While readers use past experience with outlets to form expectations (Jo, 2019), these expectations might be systematically miscalibrated, especially with clickbait stories where headlines deliberately misrepresent actual coverage. As a result, the news supply may not fully span the full support of actual consumer preferences.

Our descriptive evidence on how much time people spend consuming news in different settings offers a new approach for measuring and characterizing heterogeneity in news preferences. Inferring consumer preferences from news consumption under customization has several advantages. First, choices are made in a deliberate way that should align with people's expectations of what they are getting. For example, by choosing the *Facts Only* customization, consumers directly reveal their preference for a less opinionated article. Second, as our customization aims to vary individual attributes while minimizing "spillovers" to other dimensions, we at least partly address the issue of correlated product attributes. Third, by measuring actual consumption patterns over a prolonged period of time, our evidence is less plagued by concerns about social desirability bias compared to survey evidence (Bursztyn et al., 2025b).[19]

### 1.4.2 Experimental Design

The ideal design for measuring news preferences with customization requires four main features. First, users need a period of time to get sufficient experience with customization, actively trying different settings. Second, this learning should occur in a naturalistic, real-world news app environment rather than in a stylized lab setting. Third, to isolate the effects of customization from the content of the underlying news stories, we need a control group that reads the same underlying *distribution* of news stories without being offered customization. Finally, the design must induce exposure to customization while ensuring that all treatment groups exhibit comparable overall app use, so that any differences in downstream outcomes reflect customization rather than different usage levels or selection into the treatments.

We address these requirements by conducting a field experiment over the course of three weeks. The experiment includes four survey waves: a recruitment survey, an app download survey with randomization into treatments, a midline survey, and an endline survey. In the recruitment survey, we measure baseline characteristics and ask for the respondents' interest in getting "free early access" to an AI-tailored news app. A few days later, we invite respondents who expressed an interest in the app to a follow-up survey, where respondents who still express an interest in the app get a link to download it. Importantly, respondents only learn about their treatment status *after* they have downloaded and logged in to the app. Respondents subsequently use the app for 10 days and are paid monetary incentives for daily usage. We then administer a midline survey where we measure downstream outcomes and app perceptions. After a further 10 days, we administer the endline survey where we again elicit downstream outcomes and app perceptions. Figure 1.5 provides an overview of the experiment, which we describe in more detail below.

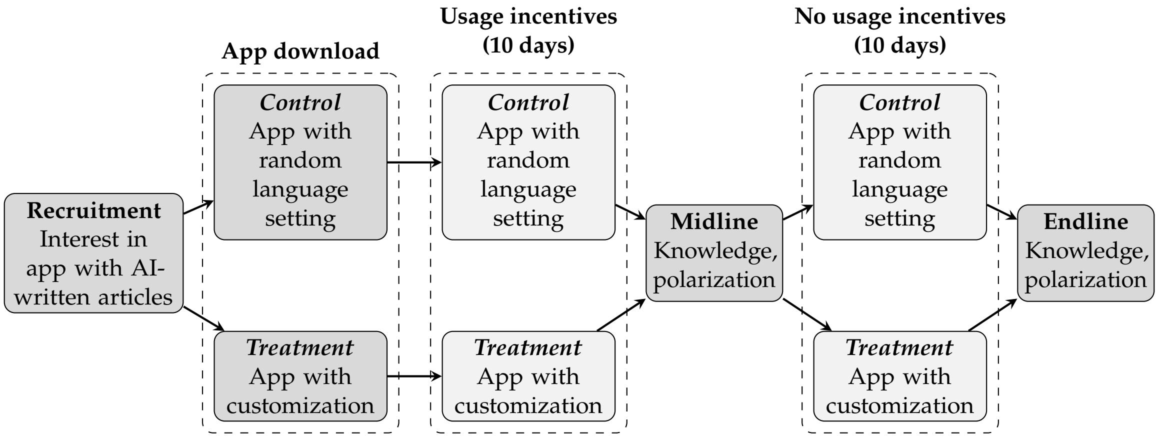
*Recruitment Survey*

The purpose of the recruitment survey is twofold. First, we want to recruit a relevant group of people interested in receiving access to an AI-based news app and who have an Android or iOS smartphone. Second, we want to measure demographics and other baseline characteristics.

The survey starts with a standard attention screener and questions about their current

[19] Although most Americans report that newspapers should not favor one side (Mitchell et al., 2018b), social desirability bias may deter readers from admitting they prefer news aligned with their views.

Figure 1.5: Overview of the Usage Incentive Experiment




*Note:* This figure presents an overview of the experiment with usage incentives to study news consumption with customization. The experiment starts with a recruitment survey. Participants who complete the recruitment survey and express interest in an app with AI-written articles are randomized into treatment (app with customization and cross-randomized to political or *Facts Only* rewrites) and control (app without customization and cross-randomized to a fixed rewrite version), and invited to the app download survey. Participants who download the app receive incentives for using the app for the next 10 days. After that, they are invited to a midline survey. Then, they can use the app for another 10 days but do not receive incentives. After that, they are invited to an endline survey.

smartphone. Respondents who fail the attention check or do not have an iPhone or Android phone are screened out of the survey. We then proceed with questions on news consumption and attitudes towards the media, including questions about which news outlets they had consumed in the prior 12 months as well as questions about news satisfaction, news fatigue, interest and trust in the news, and views on current news reporting. We also ask questions on views about AI, political attitudes (including questions on affective and issue political polarization), questions on mental health, and a rich battery of demographic questions.

To elicit interest in our AI-based news app in a natural way, we frame it as an opportunity to get "free early access" to the app before its official launch. We inform our respondents that the app offers frequent updates with relevant news articles. We highlight that the content is based on trending stories from the Associated Press and that the stories are rewritten using AI technology "to deliver essential information quickly and clearly." Finally, we emphasize that the core facts remain unchanged from the AP sources, "guaranteeing the same level of trust and credibility you expect from seasoned journalists." Importantly, we do not mention customization in this description. Respondents are told that those who express an interest in the app will be invited for a follow-up survey with instructions on how to get access.

*App Download Survey*

A few days after the recruitment survey, respondents who expressed an interest in the app are invited to the app download survey. We first ask respondents whether they are still interested in downloading the app. For those who say yes, we then give them more information about the app. Respondents are told that they will receive a $2 bonus if they download the app, enter their log-in credentials, complete a short onboarding process in the app, and enable push notifications. If they agree with this, they receive login credentials and download instructions.

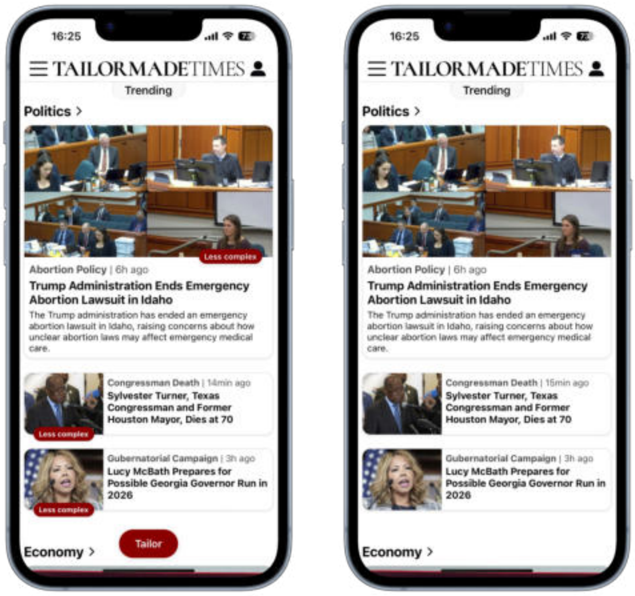
To avoid differential selection into downloading the app across treatment arms, respondents only learn about the details of the app during the onboarding process *after* they have downloaded it. Our pre-registered analysis sample includes respondents who download the app, log in, and complete the onboarding process.

**Treatment assignment** We randomly assign respondents to two different customization treatments and a control group. In both customization groups, respondents can customize news articles along four dimensions in real time. Three of the four dimensions are identical in both customization treatments, while the fourth dimension differs across treatments. Respondents in both customization treatments can choose between a regular default version or customize the articles in terms of entertainment, sentiment (positive or negative), and complexity, while holding the underlying news event constant. For the fourth dimension, we cross-randomize whether respondents can customize the political leaning of the articles (liberal or conservative) or remove opinion content from the articles. This cross-randomization allows us to make a treatment comparison between adding a political perspective or removing opinion content. As shown in Figure 1.1, respondents can, at any time, customize articles through changes in the slider settings. The order in which sliders are shown in the app is randomized between respondents.

Respondents in the control group are not offered customization and are also not aware that multiple different rewrites of the same underlying regular version exist. Instead, they are randomly assigned one of the versions for all articles they read in the app (*Regular*, *Facts Only*, *More Positive*, *More Negative*, *Less Complex*, *Liberal*, *Conservative*, or *Entertaining*). As shown in Figure 1.6, the feed looks identical for control group respondents and treatment group respondents, except that control group respondents are not shown version labels and also do not have access to the "Tailor" button. Since control group respondents are not informed about which rewrite version they receive and are also unaware that multiple versions of the articles exist, they must rely on their experience with the app to infer anything about the articles.

**App usage incentives** To provide respondents with enough experience using the app and experimenting with the different article versions and to measure the consequences of consuming customizable versus non-customizable news, it is essential to ensure high news consumption in the app and roughly equal consumption between the different treatment groups. To achieve this, we implement a series of usage incentives. First, we make the $2 bonus for downloading the app conditional on enabling push notifications. Second, after people have completed the onboarding process and enabled push notifications in the app, we inform them about a second bonus opportunity in which they earn $1 for each day they read at least three articles in the app over a 10-day period. Note that we do *not* incentivize using the customization slider. Third, we offer participants an additional $2 bonus if they keep push notifications enabled for the following 21 days. Since we observe all activities within the app, we can monitor compliance without drawing on any self-reported data.

Figure 1.6: The AI News App: Customization vs. Control Version




*Note:* This figure shows screenshots of the app in the customization and control group. The screenshot on the left displays the app as shown to respondents in the customization treatment who have actively chosen the *Less Complex* rewrite. The screenshot on the right displays the app as shown to control group respondents cross-randomized into the *Less Complex* rewrite. Labels are only shown for respondents in the customization group. Respondents in the customization group can customize using the "Tailor" button.

*Midline and Endline Surveys*

Ten days after the app download survey, respondents are invited to the midline survey. In this survey, we measure news satisfaction, self-reported news consumption outside of the app, time substitution, and general app perceptions. In addition, we measure affective political polarization, issue polarization, factual knowledge, and mental health. Ten days after the midline survey, respondents are invited to the endline survey. The endline survey follows a similar structure to the midline survey.

*Sample*

**Setting** As in the demand experiment, we recruit respondents through Prolific, a survey platform commonly used in social science research with high data quality (Peer et al., 2022). Prolific makes it easy to pay out monetary bonuses and re-contact respondents for multi-wave studies. 17,854 participants completed the recruitment survey.[20] Of these, 53% expressed

[20]We collected the data in two separate waves: one in October/November 2024 and another in January/February 2025. While we aimed to recruit 2,500 respondents in the first wave, as indicated in the pre-analysis plan, we only managed to recruit 1,439 app users. We therefore decided to conduct a second data collection in February 2025. In this second wave, we also included the *Facts Only* customization and control arms, which were not part of the initial collection. We oversampled the *Facts Only* arm in the second wave to achieve statistical power comparable to that of the other treatment comparisons. Otherwise, the waves were identical except for minor adjustments in some of the

interest in early access to the app and were invited to the app download survey. Among those invited to the download survey, 53% started the survey, and 70% of those also downloaded the app. This resulted in a final sample of 3,478 respondents who downloaded the app and were assigned to treatment (Appendix Figure A.6). To guard against bots, our experiment required participants to pass multiple attention checks, complete CAPTCHAs, and download an app on their mobile phone.[21] The full onboarding and usage-incentive instructions are reproduced in the shared appendix (Section D.3).

We assign close to 1,500 respondents (43%) to the customization treatments (21.5% in customization with *Facts Only* and 21.5% in customization with political rewrites) and close to 2,000 respondents (the remaining 57%) to the control group (who are cross-randomized in equal proportion to each of the rewrite versions, see Appendix Table A.7). Among the 3,478 respondents who download the app and are assigned to treatment, 76.5% (2,661) participated in the midline survey and 71.7% (2,495) participated in the endline survey (Appendix Figure A.6). There is no selective attrition by treatment status across survey waves (Appendix Table A.8).

**Summary statistics** Appendix Table A.9 presents summary statistics of our final sample and Appendix Table A.10 provides evidence on the integrity of randomization. The average age is 37.7 years, with respondents ranging from 18 to 78, skewing younger than the national median of 38.9 years. About 56.4% identify as female, higher than the US average of 50.8%, and 69.1% are White, broadly comparable to the national figure of 61.6%. 87.8% of respondents report having at least some college education, exceeding the national rate of 63%. Politically, the sample leans more Democratic (42.8%) than the general population (30–35%), with 33.2% identifying as Independent and 24.0% as Republican, slightly underrepresenting Republicans. Overall, as is common for online data collections, our sample is not fully representative of the broader US population, skewing younger, more female, more educated, and more digitally oriented.

**Baseline news consumption** 58% of our respondents read the news at least somewhat frequently, 90.5% primarily access news online, significantly higher than the national average of 58%, and respondents report using an average of 1.8 news apps. To examine whether our sample has a news consumption that is broadly representative of the US online population, we map each outlet they reported reading at baseline onto the ideological alignment scale developed by Bakshy et al. (2015). This broadly used measure of slant assigns each outlet a score ranging from -2 (very liberal) to +2 (very conservative), based on the average self-reported ideology of Facebook users who shared articles from that outlet. Averaging these scores across our sample yields a slightly left-of-center, unimodal distribution with a mean of -0.123 and a Democrat–Republican gap of 0.167, as shown in Appendix Figure A.7. Most respondents cluster around centrist outlets, spanning from center-left outlets like *The New York Times* to center-right outlets like *The Wall Street Journal*. This distribution closely resembles the one observed in the nationally weighted browsing panel analyzed by Guess (2021), whose data

questions due to the new presidential administration.

[21] Downloading an app on a mobile device serves as a particularly strong anti-bot measure. Unlike attention checks or CAPTCHAs, which sophisticated scripts can sometimes bypass, app installation requires access to a real device, interaction with an app store, and user authentication. This additional friction makes it highly unlikely that automated agents could participate in the study.

exhibits a comparable mean (-0.116) and spread. Taken together, these findings suggest that our sample's baseline news diet is broadly in line with the broader US online audience.

**Selection into app downloads** While our final sample exhibits a baseline news diet broadly comparable to that of the wider US online population, there could still be some selection into who chooses to download the app. We therefore examine patterns of selection into app downloads using a broad set of baseline characteristics. For this analysis, Table A.11 contrasts the full sample of about 17,000 respondents who completed the baseline survey with our final sample that ultimately installed our app using a linear probability model. Given the high statistical power of this comparison, many coefficients are expected to be statistically significant; we thus focus on those of economically meaningful magnitude. As expected, respondents who consume the news at least somewhat frequently and those who primarily consume news online are 2.5 to 4.6 percentage points more willing to download the app ($p < 0.01$, Column 1). Download probability is also increasing in the number of news apps participants have on their phones. We also find that generalized trust in AI positively predicts the chance of downloading our AI news app. Regarding demographic characteristics, individuals with college education are 4 percentage points less likely to download the app ($p < 0.01$). Reassuringly, the ideological composition of respondents' baseline news consumption is unrelated to app downloads (Column 2).

*Compliance with Usage Incentives*

In this experiment, respondents were paid monetary incentives to engage with the news app. As a first step, we examine the extent to which respondents complied with the usage incentives. Respondents across the customization and the no-customization groups, on average, spend approximately four minutes a day in the app (Appendix Figure A.8, Panels (a) and (b)).[22] This corresponds to about 43% of the average daily online news consumption in the United States, which is 9.7 minutes (Allen et al., 2020). Respondents in both groups log in on between 6 and 7 days on average during the 10-day incentive period (Appendix Figure A.8, Panels (c) and (d)).[23] Taken together, our usage data confirms that people complied with the incentives. Moreover, there is no differential app usage across treatment arms during the incentive period—as intended by our design.

### 1.4.3 How Do People Consume News with Customization?

In this subsection, we provide descriptive evidence on people's news consumption in different customization groups that allow them to express their news preferences. For this exercise, we first derive the share of time spent in each customization setting at the individual level, and then average individual time shares across all respondents, weighting the consumption shares by the amount of time spent consuming news in the app.[24] We calculate this share for all respondents

[22] Similar to the demand experiment, we infer time spent from actions (e.g., clicks, scrolls) because we do not directly observe if the app is open on a user's phone.

[23] Appendix Figure A.9 shows that there is also no differential downloading across the customization and no customization groups, confirming that the treatment assignment worked as expected.

[24] As a robustness check, we weight the consumption shares across article versions equally for all respondents irrespective of the time they spend in the app. Appendix Figure A.10 shows that our findings are robust to this

with non-zero consumption during the respective period, i.e., those without consumption are excluded. We thus characterize how much time the *average* respondent spends reading news in different customization settings.

We first analyze usage data from the app usage incentive period (days 1–10) to ensure that differences in news consumption patterns across customization groups reflect relative news preferences rather than changes in the sample composition across customization groups.

*Customization with Political Rewrites*

In the political rewrites treatment (Panel (a) of Figure 1.7), respondents spend 37% of their time in the *Regular* setting, the neutral rewrite closest to the original AP style. Despite being the default, *Regular* accounts for just over one-third of time; almost two-thirds is spent on other options. 17% of time is spent in the *Liberal* setting, followed by *Entertaining* (14%), *Less Complex* (12%), and *More Positive* (10%) settings. Interestingly, despite experimental studies showing that more negative news leads to more engagement and higher consumption rates (Robertson et al., 2023), the only customization option that is virtually never used is the *More Negative* setting (2%). The non-use of this setting suggests that the "demand" for negativity observed with non-customizable news is likely to reflect users making choices in the moment that are inconsistent with what they actually want (Kleinberg et al., 2024). By contrast, when users are aware of the article attributes and can make deliberate choices about how the news should be reported, they actively avoid negative news.

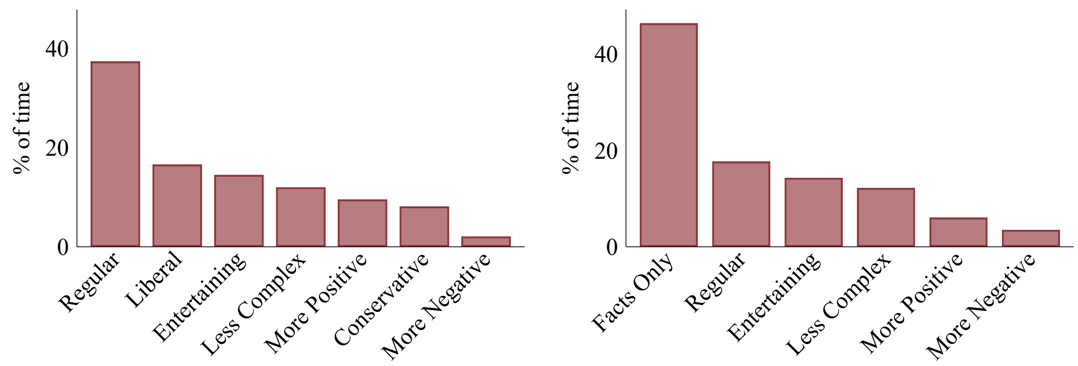
**Political heterogeneity** The *Liberal* option is used more often than the *Conservative* (17% vs. 8%), partly because the sample has more Democrats than Republicans (Table A.9). Analyzing consumption separately by party reveals strong demand for politically aligned news among both Democrats and Republicans (as shown in Figure A.11). More specifically, Democrats spend 24% of their time using the *Liberal* setting, making it second most popular after the *Regular* default among Democrats (38%). Republicans spend 25% of their time using the *Conservative* setting (compared to 39% in the *Regular* default). Democrats and Republicans are, thus, equally likely to demand ideologically aligned news. Furthermore, while the political rewrites could have been used to "see both perspectives" by switching between the *Liberal* and *Conservative* options, we find that both Democrats and Republicans use them almost exclusively to align the news with their own ideology. Our evidence is thus consistent with prior work showing that a subset of consumers have a revealed-preference for politically aligned news (Gentzkow and Shapiro, 2010). However, an important difference in our setting is that our users make a *deliberate* choice to align the news with their own ideological views. Nonetheless, the vast majority of consumers favor politically neutral reporting.

Differences in news preferences also emerge between voters and non-voters (as shown in Figure A.12). Non-voters are somewhat less likely than voters to consume the news in the default *Regular* option (29% vs. 38%) and much less likely to consume the political rewrites (8% vs 27%). Instead, non-voters spend more time reading the news in the *Entertaining* version (25% vs 13%) and *Less Complex* version (23% vs 11%). In other words, while a sizable fraction of politically active users have a demand for ideologically aligned news, non-voters who are

weighting procedure.

Figure 1.7: News Customization in the Incentive Period

(a) Customization with Political Rewrites (b) Customization with *Facts Only* Rewrites




*Note:* This figure shows the share of time spent consuming news in different customization options during the incentive period, weighting the individual consumption shares by the amount of time spent consuming news in the app. Panel (a) shows the shares for the respondents in the customization group with political rewrites ($N$ = 747). Panel (b) shows the shares for the respondents in the customization group with *Facts Only* rewrites ($N$ = 718).

disengaged from the political system instead largely prefer the news to be more entertaining and less complex.

*Customization with* Facts Only *rewrites*

Surveys and interviews suggest Americans prefer 'just the facts' reporting without opinions or commentary (Eddy et al., 2025; Mitchell et al., 2018b). However, such self-reports may reflect cheap talk or social desirability bias (Bursztyn et al., 2025b).

Consumption data from the political customization setting reveal significant demand for politically aligned news, yet the most frequently consumed option remains the politically neutral *Regular* setting. While this could be taken as evidence of a high demand for politically balanced news, a potential concern about this interpretation is that *Regular* is the default option. Thus, we cannot disentangle a demand for balanced news from default effects. Furthermore, *Regular* is not explicitly framed as a politically neutral setting.

To assess whether a large fraction of consumers actually value politically neutral and less opinionated news, we turn to our second customization treatment, which features the *Facts Only* option instead of the political rewrites option. *Facts Only* is explicitly framed to users as a neutral option that "makes the article less opinionated by focusing only on the facts," and it is also *not* the default option. This allows us to obtain a revealed-preference measure for politically neutral and non-opinionated content.

As shown in Panel (b) of Figure 1.7, the customization treatment including the *Facts Only* option demonstrates a striking and *deliberate* demand for more fact-driven and less opinionated content: Respondents spend on average 46% of their time using the *Facts Only* setting during the incentive period—a share that exceeds the time spent in any other setting by a large margin. Compared to the treatment with political rewrites, we also see a sizable decrease in the fraction of time spent in the *Regular* default option (18%) in the *Facts Only* treatment.

The time shares of the customization options that are *shared* in both customization treatments (*More Positive*, *More Negative*, *Less Complex*, and *Entertaining*) remain largely unaffected. The time share of *Facts Only* (46%) can thus almost entirely be accounted for by the time shares of *Liberal* (17%) and *Conservative* (8%) in the customization with political rewrites (which are no longer available) and the differential time use of *Regular* across the two treatment arms (37% - 18% = 19%). These patterns underscore that the high use of *Regular* in the customization treatment with political rewrites likely reflects a demand for fact-driven and neutral news rather than default effects. Furthermore, the popularity of the *Facts Only* option suggests that people's stated desire for less biased and more fact-driven news reflects a genuine preference rather than social desirability bias.

These findings raise the important question of why consumers frequently select ideologically aligned news, often presented through highly partisan opinion programs, despite our evidence indicating a substantial preference for neutral, fact-driven content. One plausible resolution is offered by theories positing that consumers perceive aligned news as inherently more accurate (Gentzkow and Shapiro, 2006). Empirical support for this explanation comes from recent studies documenting that Americans regard opinion-based programming as informative (Bursztyn et al., 2023) and perceive politically congruent news statements—even those explicitly presented as opinions—as more factual (Mitchell et al., 2018a). This phenomenon helps explain our observation of substantial demand for politically aligned content in the initial customization treatment, where the *Facts Only* option is not included, as well as the large time shares spent in the *Facts Only* rewrite.

**Heterogeneity** We only see minor differences between Republicans and Democrats in their use of the *Facts Only* option—which is by far the most popular rewrite option across both groups—demonstrating that the revealed demand for more fact-driven and less opinionated content holds across the political aisle.

One concern around our results is that they are driven by diversification motives in which respondents only prefer *Facts Only* because their existing news portfolio is politically biased. Interestingly, the demand for *Facts Only* is consistently high irrespective of whether respondents have a high or low baseline news consumption (Figure A.13). Furthermore, it is also consistently high irrespective of the slant of people's prior news consumption (Figure A.14). These results suggest that the consumption of *Facts Only* news reflects a genuine preference for more factual and less opinionated content that is not mainly driven by diversification motives.

One of the major complaints among those with low news satisfaction is that the news is too opinionated and politically biased (Reuters, 2023). If this is a genuine concern, we should expect those with low baseline news satisfaction to demand more fact-driven and less opinionated news. This is precisely what we find: Those who express greater dissatisfaction with the news at baseline are significantly more likely to use the *Facts Only* option (as shown in Figure A.15). This finding speaks to our broader motivation that news customization allows for a better match between consumer preferences and their news consumption, especially for attributes that are less covered among mainstream news outlets. The *Facts Only* option, which allows people to *transparently* avoid opinionated content, might be especially relevant in this context, as consumers have a difficult time separating facts from opinion in regular news (Mitchell et al., 2018a). Our second main result is as follows:

> ***Result.*** *We document substantial heterogeneity in news preferences. While a nontrivial share of respondents demand politically slanted news aligned with their views, the majority favor politically neutral reporting. Moreover, demand for less opinionated news is large and consistent across the political spectrum.*

*Robustness and Use of Customization Across Time and Topics*

We examine robustness across several dimensions and explore heterogeneity in customization usage. Order randomization of sliders does not influence time shares spent in the different settings (Table A.12). News preferences are consistent across topics, with similar usage distributions indicating a stable preference for particular settings (Appendix Figures A.16, A.17). After initial experimentation, users rapidly settle into stable news preferences, reflected by a sharp decline in slider adjustments and persistent usage patterns throughout the observation period (Appendix Table A.13 and Appendix Figures A.18 and A.19). Users with above-median slider use in the *Facts Only* group spend somewhat less time in the default *Regular* setting, instead consuming more news in the *Less Complex* and *More Negative* settings. By contrast, within the political customization group, we observe only minor quantitative heterogeneity based on above-median slider use (Appendix Table A.14). Finally, baseline trust in AI does not systematically affect consumption patterns under customization (Figure A.20).

### 1.4.4 Does Customization Improve Matching with Preferences?

Having documented substantial heterogeneity in individuals' use of customization, we next examine the extent to which it enhances the alignment between respondents' news consumption choices and their stated preferences for news reporting. We then investigate how consumption patterns differ between respondents who have access to customization and those in the control group, who are randomly assigned to a fixed setting.

*Perceived Mismatch and News Consumption*

To understand whether customization improves the alignment between people's news consumption and their news preferences, we exploit a series of baseline pre-treatment questions in which we ask respondents whether they think news articles are too complex, too negative, too positive, too liberal, too conservative, too opinionated, or not entertaining enough. Figure 1.8 reports results from multivariate regressions that correlate answers to these questions with the time shares spent in different settings both during the incentive period and the post-incentive period.

The results reveal that people use customization to address the perceived mismatch between their existing supply of news and their news preferences. During both the incentive and post-incentive periods, respondents who believe news articles are too opinionated spend around 10-15% of a standard deviation more time using the *Facts Only* setting ($p < 0.01$). Similarly, respondents who believe the news is too conservative spend around 30% of a standard deviation more time using the *Liberal* setting ($p < 0.01$), whereas those who think the news is too liberal spend around 10-15% of a standard deviation more time using the *Conservative* setting ($p < 0.01$). Those who think the news is too complex spend significantly more time using the *Less Complex* setting ($p < 0.01$), while those who think the news is not entertaining enough spend significantly more time using the *Entertaining* setting, though this result is only significant during the incentive period ($p < 0.01$). We also observe positive correlations between thinking the news is too negative and using the *More Positive* setting, though the point estimates are not statistically significant at conventional levels. The only dimension for which we do not observe correlations in the expected direction is the *More Negative* setting, likely reflecting that few people think the news is too positive and very few users end up using the *More Negative* setting in practice, leading to point estimates close to zero.

We can also study similar correlations between the perceived mismatch and news consumption for our control group respondents who were randomly assigned one setting for the whole period (as shown in Figure A.21). During the incentive period, when incentives induce high news consumption among all respondents, we see no systematic patterns among control group respondents. Interestingly, for the post-incentive phase, we see that respondents who believe news articles are too opinionated spend around 20% of a standard deviation more time in the app if they are assigned the *Facts Only* rewrite ($p < 0.01$). For the remaining dimensions, however, we see no systematic correlations between perceived mismatch and time use. These results suggest that customization strongly helps improve alignment with preferences relative to being assigned a setting at random. The continued predictive power of preference mismatches for control-group respondents assigned to the *Facts Only* version reinforces our earlier finding that individuals strongly prefer factual over opinion-based content. Summarizing the previous findings, our third main result is as follows:

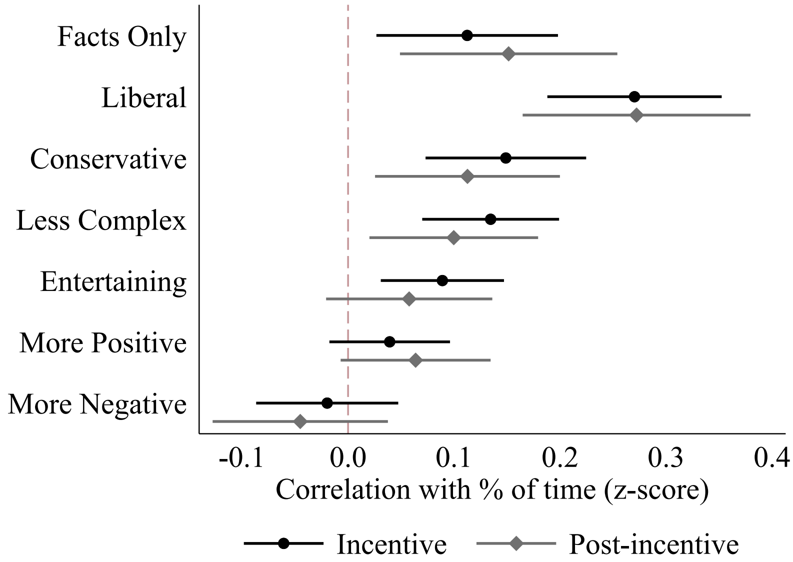
> ***Result.*** *Customization helps users address the perceived mismatch between their preferences for how the news should be presented and their actual news consumption. Those who think the existing news coverage is too opinionated, too conservative, too liberal, too complex, or not entertaining enough actively use the customization to make their news consumption more aligned with their preferences.*

*The customization wedge in news consumption*

To what extent does customization change people's news consumption towards their stated preferences, such as more fact-driven and less negative content? We answer this question by comparing news consumption in the treatment group with the control groups who were randomly assigned one version. For this exercise, we focus on the post-incentive period to capture people's news consumption in the absence of incentives once they are familiar with the app.

Differences in news consumption between the treatment and control group could emerge either through different *total* news consumption between treatments or different *time shares* spent in each rewrite version. While we observe a natural drop in consumption in the post-incentive

Figure 1.8: Perceived News Mismatch and Time Spent in Different Customizations




*Note:* This figure shows correlations between the perceived mismatch of the news and time spent in the corresponding version. Coefficients are estimated from multivariate regressions of the share of time spent with a given customization setting (e.g., *Facts Only*) on the full vector of mismatch dimensions as measured in the recruitment survey. For the mismatch dimensions, we elicit respondents' agreement with the statements "News articles are too [complex/negative/positive/liberal/conservative/opinionated]" and "News articles are not entertaining enough". We encode agreement such that larger values represent a stronger match with the customization option and standardize the values. Results are shown separately for the incentive (black circles) and post-incentive (gray diamonds) periods. All specifications include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

period (as shown in Figure A.22), there are no treatment differences in time spent on the app between respondents in the customization treatments and the control group (as further discussed in Section A.5.4 of the online appendix).[25] Any post-incentive treatment differences in news consumption thus arise from different time shares in the different settings. To benchmark news consumption patterns with customization, we therefore compare the time spent in each setting across the treatments.

To create a time share measure for control group respondents, we first scale each control arm's total time to account for differences in group sizes and then divide this by the total time of all control groups. For the treatment groups with customization, we weight the consumption shares of the two treatment groups by the amount of time spent consuming news in the app.[26] Figure 1.9 shows the consumption shares across the different article versions for respondents in the control group and compares them to the consumption shares of the treatment groups with customization. To facilitate direct comparisons between the control group and each customization treatment, Panel (a) displays outcomes for all rewrite options in the customization treatment with political rewrites, disaggregated by treatment and corresponding

[25]We also find no spillovers on self-reported news consumption outside the app (see Appendix Figure A.23).

[26]As a robustness check, we also show results for how much time the *average* user spends in the respective customization setting. Appendix Figure A.25 shows that our findings are not sensitive to how we aggregate the consumption shares.

control respondents. Panel (b) presents analogous results for the customization treatment with the *Facts Only* rewrites.

The figure reveals a striking *customization wedge* in time shares between the control groups (gray bars) and the treatments with customization (red bars). As shown in Panel (a), respondents offered customization with the political rewrites spend most of their time in the *Regular* setting (40%) and only 1% of their time using the *More Negative* setting. By contrast, the distribution among control group respondents is remarkably uniform across the different rewrite versions. Most strikingly, the *More Negative* setting has almost the same time share as the *Regular* setting. The only rewrite option that seems considerably less popular than the average among control group respondents is the *Conservative* option, likely driven by the lower fraction of conservatives than liberals in our sample. While the lower time share spent in the *Conservative* control demonstrates that news consumption can be responsive to different rewrite versions, the overall muted response across the control condition demonstrates a surprising non-responsiveness to different versions when customization is not available.

We see an even more striking customization wedge in Panel (b), where we compare respondents offered customization with the *Facts Only* option to the corresponding control groups. Respondents in the customization treatment spend about 50% of their time using the *Facts Only* option, dominating all other customization options. By contrast, the distribution among control group respondents is remarkably uniform, ranging from a time share of 12% in the *More Positive* setting to 21% in the *Facts Only* setting. All of the control group shares are close to what would be implied by an equal time spent across the settings (the dashed horizontal line in the figure).

Overall, we observe a strong difference in news consumption depending on whether they can customize. While respondents in our customization treatments overwhelmingly prefer fact-driven and less opinionated news when given the choice, our control group respondents show a remarkably uniform response to the different rewrite versions. The contrast is particularly large for *More Negative* news, which is almost never consumed among respondents offered customization but is almost as popular as *Facts Only* among control group respondents. This "customization wedge" is consistent with recent evidence of *inertia* in news consumption in which people keep consuming news misaligned with their stated preferences unless they are prompted to make an active choice about which news outlets to follow (Braghieri et al., 2026).

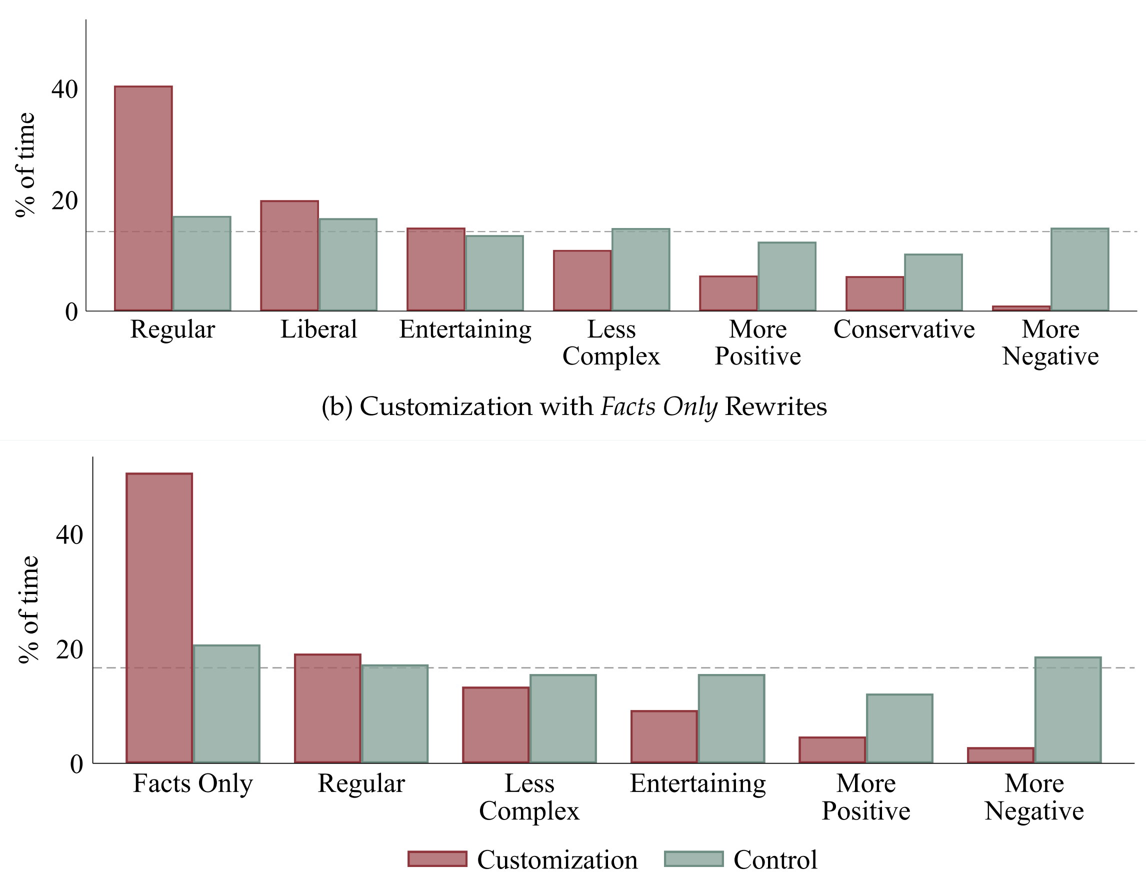
Our results demonstrate that customization allows for more sorting and a better match between underlying preferences and news consumption. Furthermore, they raise the question of whether this increased congruence leads to higher news satisfaction among respondents offered customization, which we examine in the next subsection. Given the discussion above, our fourth main result can be stated as follows:

> ***Result.*** *We uncover a substantial customization wedge in news consumption depending on whether respondents are able to customize the news. When respondents customize, they overwhelmingly seek fact-driven and less opinionated news. By contrast, respondents not offered customization spend similar amounts of time in the different versions and are much more likely to consume the negative version in particular.*

Figure 1.9: News Customization in the Post-Incentive Period

(a) Customization with Political Rewrites

(b) Customization with *Facts Only* Rewrites




*Note:* This figure shows the share of time spent using the different customization options during the post-incentive period weighted by time spent using the app. Panel (a) shows the customization treatment with political rewrites and the corresponding control arms. Panel (b) shows the customization treatment with *Facts Only* rewrites and the corresponding control arms. Panel (a) uses data from both waves, while Panel (b) uses only data from wave 2 given that we only introduced *Facts Only* rewrites in wave 2. The dashed horizontal lines show the shares that would be implied by equal time spent across settings.

### 1.4.5 Customization, News Satisfaction and App Perceptions

The previous section establishes a *customization wedge*, demonstrating that respondents offered customization consume more fact-driven and less opinionated news, whereas those not offered customization are much more likely to consume more negative news. This observation raises the question of whether the absence of customization—and thus a diminished alignment between news preferences and the provided news—also adversely affects consumer satisfaction with news consumption. Moreover, since control group respondents exhibit similar levels of news consumption across rewrite versions, evidence of reduced satisfaction in this group would suggest that observed consumption decisions absent customization poorly capture true underlying preferences.

**Empirical strategy** To examine the effects of customization on overall news satisfaction, we estimate the following pre-registered specification:

$$Y_i = \alpha + \beta\,\text{CustomizationFacts}_i + \gamma\,\text{CustomizationPol}_i + \delta X_i + \tau_t + \varepsilon_i, \tag{1.1}$$

where $Y_i$ is respondent $i$'s news satisfaction, $X_i$ is a vector of pre-treatment covariates,[27] $\tau_t$ is a fixed effect for the day on which the app download survey was completed, and $\varepsilon_i$ is the error term. We use robust standard errors for inference. Our key coefficients of interest are $\beta$ and $\gamma$, which capture, respectively, the effects of being offered customization with political rewrites and customization with *Facts Only*. The omitted group contains control group respondents who are randomized into different language settings without any customization options. $\beta$ and $\gamma$ thus capture the change in news satisfaction compared to average satisfaction in the control group.

To measure news satisfaction, we rely on the following survey item that was included in both our midline and endline collections: "How satisfied are you with the news that you have been reading over the past week?" Respondents answered this question on a five-point Likert scale ranging from 1 (*Very dissatisfied*) to 5 (*Very satisfied*). To ease the interpretation of this measure, we z-score the variable using the mean and standard deviation in the control group.

**Results on news satisfaction** As illustrated in Figure 1.10, respondents in the customization treatments report significantly higher satisfaction with news consumption compared to those in the control conditions. The customization treatment with *Facts Only* rewrites increases news satisfaction by 0.19 of a standard deviation ($p < 0.01$), while the customization treatment with political rewrites increases news satisfaction by 0.09 of a standard deviation ($p = 0.032$). Although the difference between the two customization treatments is not statistically significant at conventional levels ($p = 0.10$), the point estimates suggest that customization with *Facts Only* leads to considerably higher news satisfaction than customization with political rewrites. This is consistent with *Facts Only* being the most popular choice among the different settings.[28]

**Results on app perceptions** To better understand why customization increases news satisfaction, we next study a rich set of outcomes on perceptions of the news app. As shown in Figure 1.10, both customization treatments increase perceptions of accuracy, trust, and quality. Effect sizes on these outcomes range from 7% to 20% of a standard deviation. Consistent with the results on news satisfaction, effect sizes are consistently larger for the customization treatment with *Facts Only* rewrites. Respondents assigned customization with *Facts Only* perceive the news to be 14% of a standard deviation more fact-driven than control group respondents ($p < 0.01$), whereas respondents assigned customization with political rewrites find the news, if anything, to be somewhat *less* fact-driven. We do not observe any significant or systematic effects of customization on perceived sentiment, political bias, or entertainment.

While customization generally leads to more positive app perceptions, it also increases the perceived complexity of the app. Respondents offered customization report a roughly 20% of a

[27] $X_i$ includes demographics (age, gender, race, income, some college, full-time employed), political leaning (party affiliation and leaning, past voting), news consumption patterns (primarily consuming news online, having a paid subscription to a news platform, consuming news at least somewhat frequently, number of news apps on the phone, primary device for news consumption is mobile), trust in news and AI, and baseline news satisfaction.

[28] Figure A.26 shows the effects on satisfaction at endline.

standard deviation higher perceived complexity of the app than control group respondents not offered customization. We thus document a quality-complexity trade-off: Customization leads to higher quality perceptions but at the same time the user experience is more involved. While the net result is still higher news satisfaction on average, the perceived higher complexity of customization could partly explain why our respondents largely prefer to stick to one specific customization setting after a short initial phase of experimentation.

**Decomposing average effects** In our main specification (Equation (1.1)), we pool all control group respondents into one group irrespective of their assigned version. To better understand the drivers behind our average treatment effects on news satisfaction, we next examine disaggregated results for all control groups, making the customization treatment with political rewrites the omitted group in the regression. As shown in Figure A.27, we find that the positive average effect of customization on news satisfaction is driven by lower news satisfaction among respondents receiving the *Conservative*, *More Negative*, *More Positive*, and *Entertaining* rewrites. All of these rewrites result in news satisfaction scores approximately 20 percent of a standard deviation below those from the political customization, and roughly 30 percent below those from the *Facts Only* customization.

Interestingly, the neutral rewrites—*Facts Only*, *Regular*, and *Less Complex*—yield an overall satisfaction that is not significantly different from respondents in the customization treatments. These findings again support the notion that respondents have a latent demand for more fact-driven and politically balanced news. Furthermore, they demonstrate that consumer choices in the absence of customization are a poor measure of consumer preferences: While *More Negative* rewrites are almost never consumed with customization, they are—if anything—over-proportionally consumed among control group respondents. The fact that this consumption of negative news without customization also leads to significantly lower news satisfaction demonstrates that consumption data can lead to misleading conclusions about consumer welfare. This finding is consistent with recent evidence that overly negative or toxic content can be addictive even if it has adverse effects on news satisfaction (Beknazar-Yuzbashev et al., 2025).

**Who benefits most from fact-driven news?** A consistent finding from our experiment is that readers want *Facts Only* content: It is, by a large margin, the most popular customization choice and the customization treatment with *Facts Only* leads to an increase of 20% of a standard deviation in news satisfaction. One of our main motivations for studying customization is the high general dissatisfaction with the news media and the potential for customization to increase satisfaction through a closer alignment with underlying preferences. Given that the share of Americans closely following the news now is only 38% (down from 51% in 2016)—with potentially large negative externalities for society at large—it is particularly important from a policy perspective to increase news satisfaction among those who currently display high levels of news dissatisfaction.

To shed light on this issue, we estimate treatment effects of the customization with the *Facts Only* treatment separately for each level of baseline news satisfaction (measured on a 5-point Likert scale from 1: *Very dissatisfied* to 5: *Very satisfied*). As shown in Figure A.28, we see that the effect is largely driven by those with low baseline satisfaction. Specifically, we observe treatment effects of 23% to 34% among those who are *very* or *somewhat* dissatisfied with the news at baseline, demonstrating that customization with *Facts Only*—where respondents knowingly can select more fact-driven and less opinionated news—is an especially powerful lever for increasing news satisfaction among those who are most dissatisfied with the current news landscape for economic and political news. Given the results discussed in this section, our fifth main result follows:

> ***Result.*** *Respondents offered customization report significantly higher news satisfaction, particularly those assigned the customization treatment with the* Facts Only *option.*

### 1.4.6 Customization, Political Polarization, and Factual Knowledge

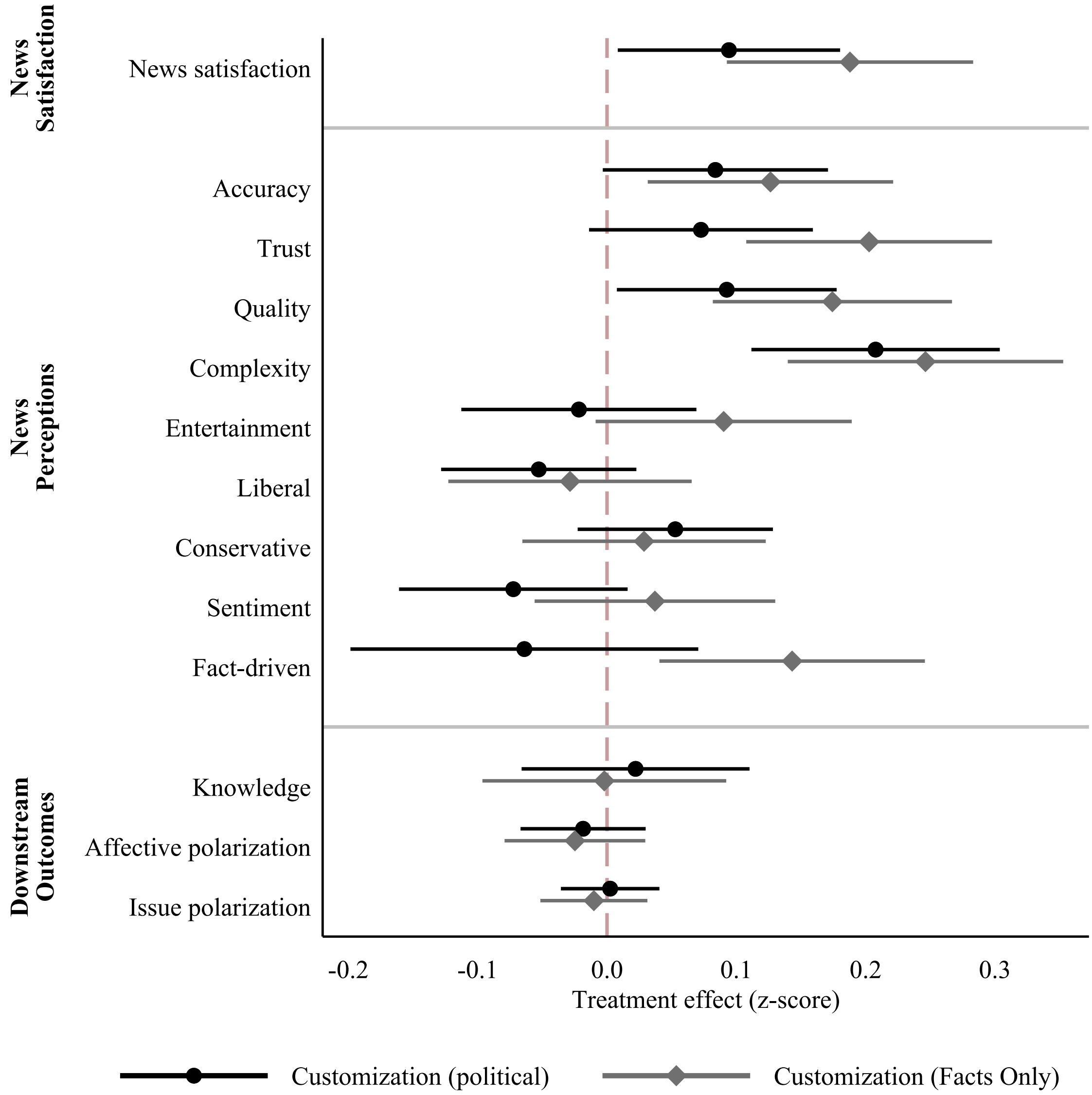
Having established that customization increases news satisfaction, we next examine its impact on political polarization and factual knowledge. Intuitively, customization involving political rewrites could have substantial effects on polarization, though the direction is theoretically ambiguous. While exposure to diverse political viewpoints within a news event may reduce polarization, the prevalent use of political customization to align content with personal ideology may exacerbate polarization by reinforcing echo chambers. Conversely, given its popularity and politically neutral framing, customization with the *Facts Only* option may reduce polarization by promoting exposure to fact-driven rather than opinionated content.

It is also unclear how customization should affect factual knowledge. On the one hand, a better match between underlying preferences and the news might increase the potential for knowledge acquisition; for instance, respondents who think the regular news is too complicated might benefit from using the *Less Complex* option. On the other hand, as discussed in Section 1.4.5, customization also increases the perceived complexity of the news, which could hinder knowledge acquisition.

**Results on affective polarization** We begin by examining *affective polarization*, defined as the extent to which individuals express more negative feelings toward those with opposing political views than toward those who share their political orientation (Iyengar et al., 2019). To measure affective polarization, we construct an index from two underlying outcomes (a standard feeling thermometer question and a hypothetical donation task). To estimate treatment effects, we rely on a specification analogous to Equation (1.1) used to estimate treatment effects on news satisfaction. As shown in Figure 1.10, we see muted treatment effects on political polarization. The point estimates are close to zero and relatively precisely estimated. Furthermore, the null effect on affective polarization holds across the different control conditions (as shown in Figure A.29), emphasizing that the overall muted effect is not driven by potentially polarizing effects in opposite directions of different control conditions.

**Results on issue polarization** We next examine results on *issue polarization*, namely the extent to which Republicans and Democrats disagree about relevant policy issues. To measure issue polarization, we construct an index based on our respondents' positions on five salient political topics: (i) free trade, (ii) police violence, (iii) sexual harassment, (iv) immigration, and (v) gun control. The index captures the extent to which individuals adopt consistently liberal or conservative positions across these domains. As shown in Figure 1.10, we also see muted

Figure 1.10: Effects on News Satisfaction, Perceptions, and Downstream Outcomes




*Note:* This figure displays the effects of customization with political rewrites and with *Facts Only* rewrites on user satisfaction and perceptions of the app, as well as downstream outcomes, relative to all control conditions. The first block shows news satisfaction. The second block shows perceptions of accuracy, trustworthiness, quality, complexity, entertainment, political bias (liberal and conservative), sentiment, and whether the content is viewed as fact-driven. The last block shows effects on news knowledge, affective polarization, and issue polarization. All effects are shown in standardized z-scores. All specifications control for demographics (age, gender, race, income, some college, full-time employed), political leaning (party affiliation and leaning, past voting), news consumption patterns (primarily consuming news online, having a paid subscription to a news platform, consuming news at least somewhat frequently, number of news apps on the phone, primary device for news consumption is mobile), and trust in news and AI. All specifications include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

treatment effects on issue polarization, with precisely estimated null effects very close to zero. This null effect also holds across the different control conditions (Figure A.29).

**Results on factual knowledge** We measure factual knowledge with a quiz of 15 questions drawn from stories featured in the app, awarding $0.20 per correct answer if completed in under 20 seconds. On average, respondents answered 8 questions correctly in the control condition. Treatment effects on factual knowledge are small and statistically insignificant (Figure 1.10), though confidence intervals are slightly wider than for polarization outcomes due to the absence of baseline controls. Our final main result follows:

> ***Result.*** *Offering respondents customization does not affect their knowledge, nor does it affect political polarization.*

**Discussion** Despite significant improvements in news satisfaction, customization has precisely estimated but negligible effects on affective polarization, issue polarization, and factual knowledge. We benchmark our treatment effects using a meta-analysis of randomized controlled trials manipulating overall news exposure or ideological content alignment. This meta-analysis reveals muted effects on polarization and factual knowledge from most interventions (details are presented in Section A.5.7 of the online appendix). Given that these studies altered overall news consumption time or ideological content alignment, their potential impact on polarization and knowledge should arguably exceed ours. Nonetheless, as shown in Figure A.31, the pooled estimates on affective and issue polarization remain close to zero, and the effects on factual knowledge are also muted and statistically insignificant despite low standard errors. In light of these results, it is not surprising that our intervention, which holds the coverage margin constant, yields muted effects on these outcomes.

## 1.5 Conclusion

We study people's news preferences using an AI-powered app that enables users to customize the characteristics of news articles while holding the news event fixed. This design overcomes a core identification challenge: Because news articles are multi-dimensional experience goods and news outlets optimize for short-term engagement rather than long-term satisfaction, market data cannot reliably reveal consumers' underlying preferences (Kleinberg et al., 2024). By allowing respondents to customize article characteristics while holding the underlying news event constant, we observe choices that directly reveal news preferences for specific attributes, such as less complex or more fact-driven content. Furthermore, as customization increases the available product space, we can observe preferences over choices not offered in the existing news market.

Using several large-scale experiments, we document six main findings. First, there is substantial demand for customization, suggesting a mismatch between people's preferences and the existing supply of news that they think customization can help address. Second, we observe substantial heterogeneity in how people use customization. While a significant minority demand politically aligned news, the large majority demand more fact-driven and less opinionated news. Third, customization helps users address the perceived mismatch between their preferences for how news should be presented and their actual news consumption. Fourth, we uncover a "customization wedge" in which respondents not offered customization end up consuming news misaligned with their stated preferences. Fifth, customization leads to significantly higher news satisfaction, particularly when it includes the option to make the news more fact-driven and less opinionated. Finally, while customization increases news satisfaction, it has no significant effects on political polarization or factual knowledge—consistent with evidence from other studies showing that media interventions more generally have muted impacts on these outcomes.

Given the inherent challenges in identifying news preferences from existing market data, tools that allow consumers to make active and deliberate choices about how news is presented are essential for understanding people's underlying preferences. Our results that most people prefer fact-driven news without opinions are consistent with the idea that the observed demand for slanted news in existing media markets to a large extent reflects quality considerations, as in Gentzkow and Shapiro (2006), rather than a strict preference for ideologically aligned content. Consistent with this interpretation, Braghieri et al. (2026) find that information about bias makes people follow less slanted news outlets on social media.

More broadly, our intervention not only advances understanding of news preferences but also informs policy discussions on the implications of customization as a platform feature. A natural question is why leading news outlets have not already implemented customization more broadly. Given that real-time news customization is only made possible by recent advances in Generative AI, the lack of mainstream adoption could reflect a natural time-lag in adopting new technology. However, it is also worth asking whether it would be profitable for news outlets to offer news customization. While classic models of product differentiation suggest that catering to heterogeneous consumer preferences should increase profitability, it is not obvious that this is the case in the market for news, where many news outlets optimize for engagement rather than consumer satisfaction. Given the importance of social media networks in driving the traffic of many news outlets, it might still be optimal from a revenue perspective to supply "clickbait stories" rather than to offer people more control through customization. Furthermore, while customization increases news satisfaction, it also makes the news seem more complex, potentially further limiting its appeal to news outlets. More work is needed to understand these dynamics further.

# Chapter 2

# The Asymmetric Response to Political Spin

*joint with Felix Chopra, Ingar Haaland, Christopher Roth, and Vanessa Sticher*

**Abstract:** We study how political spin – the strategic presentation of facts to advance a particular political perspective – shapes news consumption and political polarization. Using a custom-built AI news app, we conduct a field experiment in which respondents are randomly assigned to receive conservative, liberal, or politically neutral versions of the same news articles, isolating variation in spin while holding constant the underlying news events. Our results reveal a strongly asymmetric response to political spin. Once app usage incentives are removed, we find that misaligned spin reduces news consumption, while aligned spin does not increase it. Consistent with these patterns, perceived accuracy and news satisfaction decline when respondents are exposed to politically misaligned spin but do not improve under aligned spin. Turning to downstream consequences, we find that aligned spin modestly increases issue polarization, while misaligned spin has no detectable effects on issue or affective polarization. These findings challenge demand-side models predicting that consumers reward politically aligned reporting. Instead, they suggest that consumers punish misaligned framing without rewarding aligned framing, implying that the persistence of political spin is more likely driven by supply-side forces.

*Keywords:* Spin, News Consumption, Market For News, Media Bias, Polarization

*JEL codes:* C93, D83, L82, P00

## 2.1 Introduction

When audiences seek out news that echoes their own convictions, media outlets have a strong incentive to frame the same facts in ideologically aligned ways. Demand-side models of media bias predict precisely this equilibrium: readers reward like-minded reporting, so profit-maximizing firms slant their language to match partisan tastes (Gentzkow and Shapiro, 2006; Mullainathan and Shleifer, 2005). While such alignment may satisfy consumer preferences, it also raises broader concerns about the societal costs of political spin. A growing literature warns that ideologically slanted language can entrench partisan divisions and foster echo chambers (Durante et al., 2019; Levy, 2021), potentially undermining democratic discourse (Sunstein, 2018). Yet isolating the causal effects of spin is difficult, because real-world partisan media typically bundle framing together with differences in story selection and coverage.

This paper studies how the political spin of news — the strategic presentation of factual information to promote a particular political perspective — affects perceived article quality, news consumption, and political polarization. To answer this question, we built an AI news app, which allows us to systematically vary the political spin applied to identical news events and to measure people's actual news consumption. Our articles are generated by an AI-based system using content sourced from a highly credible outlet, the Associated Press. In particular, we develop and validate an automated procedure leveraging recent advances in large language models to produce news articles with either liberal, conservative or neutral spin. Our experimental design isolates variation in political spin while holding constant the coverage of news events, allowing us to identify the causal effects of spin on actual news consumption and its downstream effects on perceptions and attitudes.

We study these questions in a field experiment with more than 1,000 Americans recruited on Prolific. Respondents download and use our AI news app and are randomly assigned to receive politically neutral, liberal, or conservative versions of the articles.[1] In our main analysis, we map exposure into *aligned* and *misaligned* spin using respondents' partisan identification. We incentivize app use during an initial 10-day period so as to equalize exposure across treatment arms and make downstream comparisons on perceptions, knowledge, and polarization more interpretable. An important feature of the design is that respondents know they are using an AI-powered news app, but they are not informed that different political rewrites of the same stories exist across users. Any updating about slant, credibility, or quality must therefore come from the reading experience itself rather than from treatment labels or prior beliefs about outlet identity.

Our design maps naturally into a simple empirical test of demand-side theories of politically aligned news. If consumers symmetrically reward aligned framing, then politically *aligned* spin should increase engagement and satisfaction relative to the neutral control, while *misaligned* spin should reduce them. By contrast, if the main response margin is aversion to counter-attitudinal framing rather than active demand for aligned spin, then we should observe a pronounced penalty for misaligned spin but little or no premium for aligned spin. This contrast is central to how we interpret the treatment effects below.

[1] A companion paper analyzes additional groups in which users customize their news content in terms of sentiment, entertainment, opinion content, political leaning and complexity in real time (Chopra et al., 2025).

We report three main findings. First, political spin has strongly asymmetric effects on news consumption. During the 10-day incentive period, consumption levels are similar across the neutral, aligned, and misaligned groups. Once these incentives are removed, however, respondents exposed to misaligned spin engage less with the app: they log in on fewer days and spend 0.13 SD less time in the app per day ($p < 0.01$). For time spent, this corresponds to about 0.16 fewer minutes per day, or about 1.6 fewer minutes over the 10-day post-incentive period. By contrast, aligned spin does not increase news consumption relative to politically neutral reporting. The estimated effect on daily time spent is small at 0.02 SD and statistically indistinguishable from zero.

Second, this asymmetry is mirrored in how respondents evaluate the content in the app. Respondents clearly recognize the direction of the framing: perceived political alignment rises by 0.42 SD under aligned spin and falls by 0.69 SD under misaligned spin. But only misaligned spin causes respondents to downgrade the content. Relative to the neutral benchmark, misaligned spin lowers perceived accuracy by 0.38 SD, trustworthiness by 0.39 SD, quality by 0.19 SD, and news satisfaction by 0.33 SD. Aligned spin, in contrast, does not improve perceived accuracy, trustworthiness, or satisfaction relative to neutral reporting. In other words, respondents punish misaligned framing, but do not reward aligned framing with higher perceived credibility or satisfaction. Heterogeneity analyses suggest that this pattern is driven primarily by respondents with stronger ideological commitments, for whom politically misaligned framing sharply reduces both perceived credibility and subsequent engagement.

Third, the downstream consequences for learning and attitudes are limited overall, but not entirely absent. We find no meaningful effects on factual knowledge or affective polarization. At the same time, aligned spin leads to a modest increase in issue polarization of 0.09 SD that survives adjustment across the paper's three headline downstream outcomes, while misaligned spin has no detectable effect. The component evidence of the issue polarization result points most clearly to trade-related attitudes rather than to a uniform shift across all issue domains, so we interpret this result cautiously. However, it suggests that ideologically aligned framing can reinforce some policy divisions even when the underlying news events are held fixed.

Taken together, the findings place limits on simple demand-side accounts in which politically aligned spin is rewarded symmetrically by consumers. In our setting, aligned spin does not increase engagement or satisfaction, whereas misaligned spin substantially reduces both. More broadly, the paper isolates an important and understudied dimension of media bias: the effect of spin holding fact selection fixed. This distinction matters for how we think about media markets and policy. If the main effects of partisan media arise from selective coverage, as in Broockman and Kalla (2025), then the relevant margin is difficult to discipline normatively. If framing itself materially changes behavior and beliefs, then the case for neutral journalistic standards becomes clearer.

**Literature** Our study contributes to a large literature on the market for news (Demsetz and Lehn, 1985; Mullainathan and Shleifer, 2005). This literature has documented a high degree of media bias using a variety of different measurement approaches (Bakshy et al., 2015; Cagé et al., 2025; Gentzkow, 2018; Gentzkow and Shapiro, 2010; Gentzkow et al., 2014; González-Bailón et al., 2023; Green et al., 2025; Groseclose and Milyo, 2005; Qin et al., 2018). This line of research has analyzed both demand-side and supply-side forces that shape the structure and functioning of the news market (Anderson and McLaren, 2012; Baron, 2006; Durante and Knight, 2012; Gentzkow and Shapiro, 2006; Prat and Strömberg, 2013). We contribute to this literature by providing field-experimental evidence that consumers punish misaligned political framing but do not symmetrically reward aligned framing.

Our paper also contributes to a literature examining the effects of mass media on political attitudes and polarization (Allcott et al., 2024; Broockman and Kalla, 2025; DellaVigna and Kaplan, 2007; Gerber et al., 2009; Guess et al., 2023b; Martin and Yurukoglu, 2017). Existing work shows that exposure to different news environments can shape learning, attitudes, and partisan animus, but the underlying treatments often combine multiple margins at once. For example, Gerber et al. (2009) randomize subscriptions to the Washington Post and Washington Times, and document muted effects on opinions. Broockman and Kalla (2025) show that incentivizing Fox News viewers to watch CNN leads to learning and attitude moderation. Levy (2021) finds that exposure to counter-attitudinal news reduces negative affect toward the opposing political party but does not shift political beliefs. In a large-scale field experiment during the 2020 U.S. presidential election, Nyhan et al. (2023) reduced Facebook users' exposure to like-minded content by one-third, increasing cross-cutting exposure and reducing uncivil language, but found no detectable effects on polarization or political attitudes, suggesting limited causal impact of echo chambers despite their prevalence. Djourelova (2023) shows that reduced exposure to slanted language lowers support for restrictive immigration policies. We complement this literature by showing that framing alone has limited downstream effects overall, with the clearest effect being a modest increase in issue polarization under aligned spin.

Finally, we build on a growing body of work that investigates individuals' preferences over news content (Capozza et al., 2026; Thaler, 2024). Chopra et al. (2024) conduct a survey experiment to study how beliefs about filtering bias shape the demand for news. Chopra et al. (2022) study the effect of fact-checking on subscriptions to newsletters. Chopra et al. (2025) characterize how people consume news when they can deliberately customize several features of news articles in real time, e.g. by making the articles more entertaining, less complex, less opinionated or more positive. They uncover that customization increases the share of readers that consume the news in a version that removes opinion content. In a large-scale Facebook field experiment, Braghieri et al. (2026) show that behavioral and informational frictions distort news consumption, but that a simple interface for re-optimizing followed pages leads users to adopt a more balanced and reliable news diet. Our findings indicate that readers are more responsive to counter-attitudinal framing than to aligned framing, consistent with punishment of misaligned spin rather than strong positive demand for aligned spin.

Methodologically, this paper contributes to a growing literature conducting media experiments with actual news consumption data (Allcott et al., 2022; Beknazar-Yuzbashev et al., 2025; Braghieri et al., 2026; Jo, 2019; Levy, 2021) and to a literature using large language models to analyze media markets (Braghieri et al., 2024). Our key methodological contribution lies in designing and running our own AI-powered news app that delivers real-time coverage of the same news events rewritten to reflect different political spin. This design enables us to experimentally isolate the effects of political spin while holding the coverage of news events and the filtering of facts constant.

This paper proceeds as follows: In Section 2.2, we describe the AI news app. In Section 2.3, we describe the experimental design. Section 2.4 presents the main experimental results on perceptions, news consumption and political polarization. Section 2.5 discusses external validity. Section 2.6 concludes.

## 2.2 The AI News App

In this section, we introduce TailorMadeTimes (TMT), the AI-powered news app we developed to conduct large-scale experiments to study news consumption. Using our own mobile app has two central advantages for the purposes of this paper. First, it allows us to observe respondents' actual reading behavior in a naturalistic environment, making our estimated treatment effects more externally valid and less prone to social desirability bias (Bursztyn et al., 2025b) or experimenter demand effects (de Quidt et al., 2018). Second, it lets us vary the framing of the same underlying news event while holding the set of stories shown to users fixed. This design is important for identification: differences across treatment arms arise from differences in presentation and rhetorical emphasis rather than from differences in which events are covered.

### 2.2.1 News Sourcing and Feed Curation

TMT is a mobile news app designed to resemble standard news apps while giving us control over how articles are presented. The home feed is organized into six sections—"Politics", "Economy", "U.S.", "World", "Lifestyle", and "Sports"—and displays the top three trending articles in each section. Article previews include a title, timestamp, short topic label, and, for most stories, an image; for the top article in a section, the lead is also shown directly in the feed. Users can click into dedicated section pages with additional articles, and older articles remain accessible through search and bookmark features. The app refreshes hourly, which results in a continuously updating feed with newly available articles throughout the day. Users who enable push notifications receive four notifications per day, each featuring the most trending story since the previous notification.

All articles shown in the app are based on reporting from the Associated Press (AP), which is widely regarded as a highly credible and comparatively nonpartisan news source. We retrieve English-language AP articles via a news API subscription (www.newsapi.ai) and restrict attention to articles between 100 and 1,000 words. For each article, we construct a trending score that combines salience and recency. Salience is based on how often entities mentioned in the article—such as people, organizations, or locations—appear across all indexed news sources over the preceding two days. We combine this measure with recency and rank articles accordingly, such that more salient and more recent stories appear higher in the feed. Importantly, this ranking is identical for all users: treatment assignment changes how a given story is written, not which stories are shown or how they are ordered in the app.

For this experiment, the relevant sections of the app are "Politics", "Economy", "U.S.", and "World". These are the sections in which we deploy liberal and conservative rewrites. By contrast, "Lifestyle" and "Sports" offer only limited scope for meaningful political framing and therefore remain in politically neutral versions throughout the experiment.

### 2.2.2 Article Versions

The baseline version in TMT is the *Regular* version, which is a neutral summary of roughly 300 words based on the original AP article. Building on this baseline, we generate alternative versions that vary the article's presentation while keeping the underlying news event fixed. The key treatment dimensions are political: *Liberal* rewrites frame the same event in language that is more favorable to Democratic positions or political actors, while *Conservative* rewrites use language that is more favorable to Republican positions or political actors. These rewrites operate through differences in framing, emphasis, and rhetorical choice rather than by changing the set of facts covered.

**Political Spin** We define political spin as the strategic presentation of factual information in a way that emphasizes certain aspects over others to align with or promote a particular political perspective. Motivated by the seminal analysis of partisan media framing in Gentzkow and Shapiro (2010), we develop AI agents that systematically vary the political spin applied to a given news story. The agents generate conservative and liberal versions of each article by adjusting language to portray the Republican and Democratic parties, respectively, in a more favorable light. For instance, conservative rewrites may refer to the "death tax," while liberal versions prefer the term "estate tax." These adjustments shift rhetorical emphasis while holding the underlying news events constant and avoiding unsubstantiated causal claims.

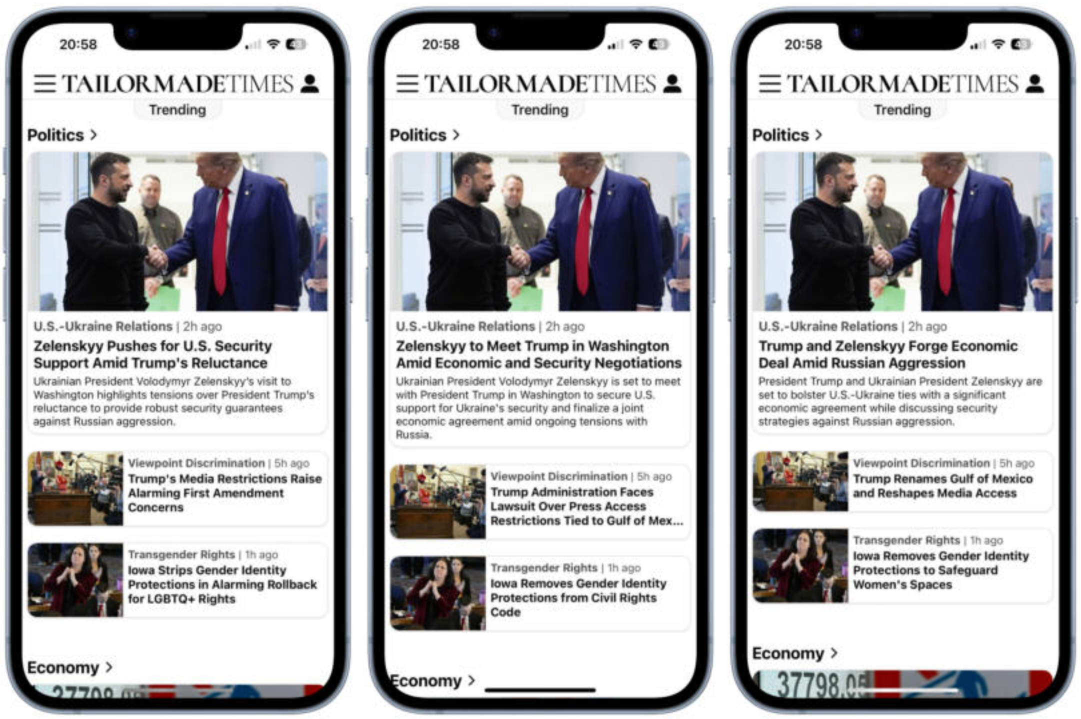
Figure 2.1 illustrates how political spin is operationalized in our experiment through systematic variation in the language and framing of news events. Each app version displays the same set of underlying news stories, but with adjustments to tone, emphasis, and evaluative content that reflect liberal, neutral, or conservative spin.

The headline article on U.S.-Ukraine relations provides a clear illustration. In the *liberal* version (left panel), the story emphasizes President Zelenskyy's active push for U.S. security support despite Trump's reluctance, highlighting "tensions" and framing Trump's stance as hesitant or resistant. In contrast, the *conservative* version (right panel) presents the narrative around forging an "economic deal" between Trump and Zelenskyy, suggesting proactive and positive diplomatic action rather than reluctance. The *neutral* version (center panel) adopts a balanced descriptive tone, presenting negotiations on security without evaluative cues, merely noting Zelenskyy's visit and the potential deal.

Framing differences are also evident in subsequent articles. On viewpoint discrimination, the liberal version highlights alarming "First Amendment concerns" raised by Trump's media restrictions. In contrast, the conservative version frames the situation around Trump actively reshaping media access, implicitly suggesting strength and decisiveness rather than restriction or concern. Similarly, the coverage of transgender rights differs notably. The liberal framing focuses on the "alarming rollback" of LGBTQ+ rights protections, signaling progressive advocacy, whereas the conservative version stresses the removal of protections as safeguarding "women's spaces," aligning with traditional conservative values. The neutral version presents these events straightforwardly, avoiding evaluative language.[2]

**Politically neutral versions** In addition to these political versions, the broader TMT system also generates politically neutral rewrites. The politically neutral versions relevant for this paper are *Regular*, *Less Complex*, which simplifies language and sentence structure using guidelines similar to *Simple English Wikipedia*, and *Facts Only*, which removes evaluative and opinionated language from the *Regular* version while preserving factual content. In the main analyses below, we pool the *Regular*, *Less Complex*, and *Facts Only* versions into a politically neutral control group in order to maximize statistical power. We view this pooled control as appropriate for the present comparison because the three versions are all close to neutral on the political dimension of interest, exhibit similar consumption patterns and similar core perception results in Appendix Section B.12, and remain close to AP in the spin analysis below (see in particular Appendix Figure B.4). The benchmark in the paper is therefore a pooled AP-based neutral environment rather than any single neutral rewrite or any existing partisan outlet.

Figure 2.1: TMT App




*Note:* These screenshots show the home feed of the TMT app taken on February 28, 2025 (during wave 2 of the experiment). The left panel shows the app with a *liberal* spin, the middle panel shows the politically *neutral* version based on the regular rewrite, and the right panel shows the app with a *conservative* spin. Figure B.1 presents the three politically neutral versions—*Regular*, *Less Complex*, and *Facts Only*—separately.

### 2.2.3 Rewrite Pipeline and Safeguards

To generate these article versions systematically, we use a multi-agent rewriting system built on OpenAI's GPT-4o model family. The objective of this system is to vary framing while preserving

[2]See Appendix Section B.1 for titles, leads and full bodies of all rewrites for the first story.

the same underlying news event and maintaining journalistic integrity. The pipeline has three stages: preprocessing, writing, and editing.

In the preprocessing stage, an agent determines whether a given rewrite dimension is appropriate for a given article. This matters because some article-version combinations should not be generated. For example, political rewrites are not applied to “Sports” or “Lifestyle” stories, and other dimensions in the broader TMT system are restricted for sensitive content. This stage ensures that the rewrite system operates only where the requested variation is substantively meaningful and appropriate.

In the writing stage, an agent first generates the *Regular* version as a concise summary of the AP source article. Conditional on the article and the set of eligible versions, specialized rewriting agents then produce the alternative versions. These agents are instructed to preserve the underlying event, maintain coherence and article length, and avoid misleading, omitting, or fabricating facts. They are also given version-specific instructions and curated few-shot examples that teach the desired framing style; the shared and version-specific prompts are documented in the shared prompts appendix (Sections E.1 to E.4, E.7 and E.9). For the political versions, the relevant variation concerns framing and rhetorical emphasis, not the introduction of new factual claims or changes in coverage.

In the editing stage, an editor agent produces version-specific titles, leads, and push notification text. Finally, all generated content is screened using OpenAI’s Moderation API. If content is flagged, or if an article receives the lowest appropriateness score in preprocessing, the article is withheld from automatic publication and sent to us for review. Unflagged articles are published automatically. This combination of version-specific prompting, preprocessing restrictions, and moderation safeguards is designed to ensure that observed treatment effects reflect differences in political spin rather than obvious errors in the generated content.

#### 2.2.4 Validation of Rewrites

The interpretation of our experiment depends on two conditions: first, the political rewrites must preserve the same underlying news event; second, they must create meaningful and perceptible differences in political spin. We therefore validate the rewrite system using both a fact-level consistency procedure and a human validation exercise. The companion paper (Chopra et al., 2025) describes the broader TMT rewriting system in more detail; here we summarize the validation evidence most relevant for this paper.

**Fact consistency** To assess whether rewrites preserve the underlying event, we use a custom fact-level alignment procedure. Starting from each article version, we extract factual claims and classify them into *core* and *supplementary* facts. Core facts are those necessary to reconstruct the article’s main event (who did what and why it matters) whereas supplementary facts contain background information, stylistic additions, or explanatory detail. We then compare the core facts in the *Regular* version with those in each rewrite in both directions. Operationally, we first align facts using embeddings, calibrating the cosine-similarity threshold on a held-out set of human judgments, and then use an LLM adjudication step to re-evaluate unmatched cases. As an additional check, we verify numeric consistency separately for dates, quantities, and percentages. Appendix Section B.2.1 reports the full procedure and results.

The resulting evidence shows very high alignment for the versions relevant to this paper. After LLM adjudication, the share of *Regular* core facts covered by the rewrite is 96.8% for *Conservative*, 98.0% for *Liberal*, 98.2% for *Less Complex*, and 96.2% for *Facts Only*; the corresponding reverse-coverage rates are 95.9%, 95.8%, 99.4%, and 95.8%, respectively (Appendix Table B.1). Numeric alignment is close to 100% for all these versions. Taken together, this evidence indicates that the rewrites preserve the same underlying news event to a very high degree.

**Human validation** We complement the fact-level exercise with a human validation study in which respondents compare the *Regular* version of an article to one rewritten version and indicate which one is more conservative, more liberal, less complex, or more fact-driven. Figure B.2 reports the results for the versions most relevant here. Respondents correctly identify the intended treatment dimension at high rates: in the validation study, raters correctly classify the political rewrites and the *Facts Only* rewrite in roughly 70% of cases, and several other rewrite dimensions in roughly 80% of cases. Spillovers into other dimensions are modest overall. In particular, respondents reliably perceive the *Liberal* and *Conservative* rewrites as more liberal and more conservative, respectively. They also perceive *Facts Only* as more fact-driven and *Less Complex* as less complex, which supports our use of these versions as politically neutral controls rather than as latent political treatments. Appendix Section B.2.2 provides the corresponding figure and additional detail.

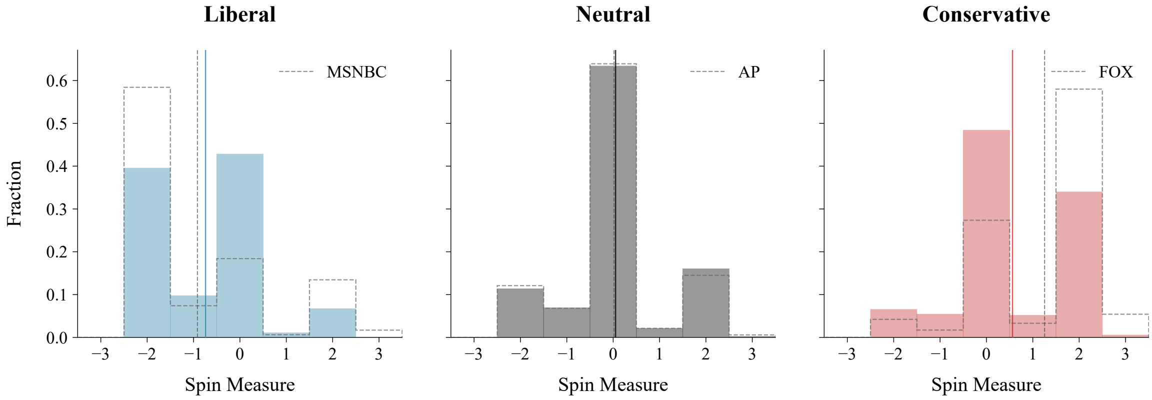
### 2.2.5 Spin Analysis of Article Versions

Having established that the rewrites are perceived as intended and preserve the same underlying stories, we next quantify how much ideological movement they generate. To do so, we classify article versions using the method of Braghieri et al. (2024), which conceptualizes media bias as arising from partisan language, explicit political positioning, and coverage choices. Because coverage is held fixed by construction in our setting, we adapt their procedure to focus on the first two components only: partisan language in the spirit of Gentzkow and Shapiro (2010) and explicit political positioning within the article text.

Operationally, we classify each article using GPT-4o with the Braghieri et al. (2024) prompt, modified only to remove the coverage component. We apply this procedure to all rewritten articles in the four politically relevant app categories—"Politics", "Economy", "U.S.", and "World"—which yields 6,463 unique underlying news stories, each observed in multiple rewritten versions. We transform the original 1–7 scale into a centered spin measure ranging from $-3$ to $+3$, where negative values correspond to content more favorable to the Democratic party, positive values correspond to content more favorable to the Republican party, and zero corresponds to political neutrality.

Figure 2.2 presents the distribution of this measure for the liberal, neutral, and conservative versions. The liberal distribution is shifted substantially to the left relative to the neutral versions, whereas the conservative distribution is shifted to the right. Across all articles, the mean spin measure is $-0.74$ for *Liberal* articles, 0.05 for the pooled neutral versions, and 0.56 for *Conservative* articles. Relative to the cross-article standard deviation of the spin measure, these correspond to shifts of about 0.65 SD to the left for liberal rewrites and 0.42 SD to the right for conservative rewrites, both highly statistically significant ($p < 0.001$). Appendix Figure B.3 shows that the same directional pattern appears within each news category.

Figure 2.2: Distribution of Spin




*Note:* This figure shows the distribution of spin for the 6,463 unique underlying news stories in our study period within the relevant app categories ("Politics", "Economy", "U.S.", and "World"), separately for the *liberal*, *politically neutral*, and *conservative* versions. The scale goes from −3 (extremely favorable to the Democratic party) to +3 (extremely favorable to the Republican party). The spin measure is based on the methodology of Braghieri et al. (2024). The dashed lines show spin for MSNBC, Associated Press, and Fox articles using the same measure for the same time frame and categories, standardized to the category composition of our sample. The vertical lines show the means of the distributions.

The figure also benchmarks our rewrites against real-world news content. Associated Press lies close to the center, MSNBC is clearly to the left, and Fox is clearly to the right. This comparison serves two purposes. First, it shows that the political rewrites move content in ideologically meaningful directions that resemble real outlet differences. Second, it shows that the politically neutral versions remain close to the underlying AP articles. In particular, the *Regular* and *Less Complex* versions are statistically indistinguishable from AP in mean spin. The pooled neutral distribution is slightly right-shifted because the *Facts Only* version is somewhat more conservative on this measure, but the magnitude is substantively very small—about 0.03 SD relative to AP. We therefore interpret the pooled control as a high-quality neutral benchmark for the present comparison, with the appendix providing additional validation for this pooling choice (Appendix Figure B.4 and Appendix Section B.12).

**Drivers of spin** To understand what drives these shifts in measured ideology, we decompose the change in spin from the *Regular* version to the political rewrites into phrase-level contributions. Following the logic of Gentzkow and Shapiro (2010), we estimate ideological weights for two- and three-word phrases using differences in language between conservative and liberal rewrites and then combine these weights with changes in phrase usage relative to the *Regular* version. We group related phrases into broader categories and aggregate their contributions to the overall shift in the article-level spin measure.

Table 2.1 shows that conservative spin is driven primarily by categories such as criticism and contestation, fiscal and taxpayer language, social and moral values, security and order, and partisan positioning. Liberal spin, by contrast, is driven mainly by urgency and crisis framing, references to political actors, emphasis and salience, public investment and services, and equity and rights. Appendix Table B.2 provides the underlying phrase-level examples. This decomposition helps establish that the rewrites operate through interpretable framing choices rather than through latent changes in topic selection.

Table 2.1: Drivers of Spin in Political Rewrites

| Conservative spin | | Liberal spin | |
|---|---|---|---|
| Category | Share (%) | Category | Share (%) |
| Criticism and contestation | 24.0 | Urgency and crisis | 25.8 |
| Fiscal and taxpayer | 20.7 | Political actors | 22.9 |
| Social and moral values | 10.1 | Emphasis and salience | 19.2 |
| Security and order | 8.3 | Public investment and services | 10.1 |
| Partisan positioning | 8.3 | Equity and rights | 5.1 |

*Note:* This table decomposes the mean change in the spin index induced by political rewrites into phrasing categories. Categories are constructed by grouping short phrase families (2–3 word n-grams), and category contributions are obtained by aggregating phrase-level contributions to the mean shift in the spin index. Shares are percentages of total positive category contribution on each side. The table reports the five largest labeled categories separately for conservative and liberal rewrites relative to the corresponding regular version. Appendix Table B.2 reports the corresponding phrase-level examples.

Overall, the evidence in Figure 2.2 and Table 2.1 shows that the rewrite system generates substantial, interpretable, and directionally correct variation in political spin while keeping the underlying stories fixed. This is the key methodological input for the experimental analysis below: when we compare aligned, misaligned, and politically neutral versions, we are isolating the effects of political framing rather than differences in coverage.

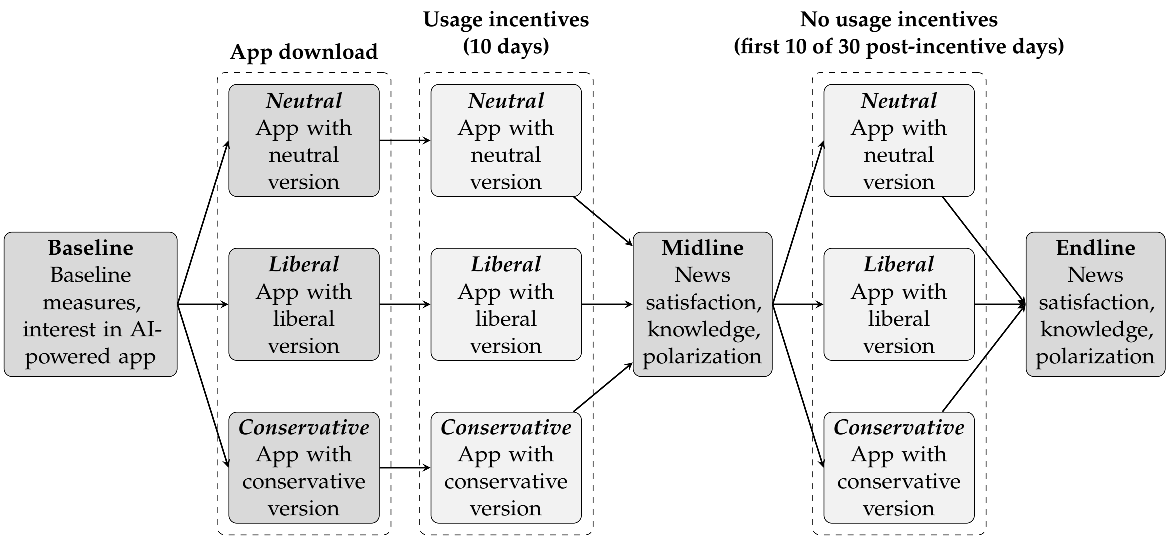
## 2.3 Experimental Design

### 2.3.1 Overview

We recruit smartphone users through an online survey, screen for interest in an AI-powered news app, and collect baseline measures of demographics, media use, and political attitudes. In our main analysis, we then use respondents' partisan identification to reclassify exposure to the liberal and conservative rewrites as either *aligned* or *misaligned* political spin, while the pooled politically neutral versions serve as the control group. Spin is varied through linguistic framing while holding the underlying news events constant. To equalize initial usage of the app across conditions, we implement financial incentives for app engagement during the first 10 days after onboarding. Midline and endline surveys are fielded 10 and 20 days after the app download survey, respectively, while app-based consumption is observed for a total of 40 days: the 10-day incentive period plus three subsequent 10-day post-incentive periods. All in-app activity is passively tracked to avoid reliance on self-reports. Figure 2.3 provides an overview of the experiment. The full participant instructions for the app-based experiment are reproduced in the shared appendix (Section D.3).

At the assignment level relevant for this paper, respondents are allocated to one of five article environments: *Regular*, *Less Complex*, *Facts Only*, *Liberal*, or *Conservative*. For the main estimands, we then group the first three versions into a pooled politically neutral control condition and recode the liberal and conservative assignments using respondents' partisan identification: a liberal rewrite is *aligned* for Democrats and *misaligned* for Republicans, while a conservative rewrite is *aligned* for Republicans and *misaligned* for Democrats. Respondents who do not identify with either party and do not report a partisan lean are excluded from these aligned-versus-misaligned specifications. As a result, the main treatment effects compare aligned and misaligned political spin to the neutral control group.

Figure 2.3: Experiment Overview




*Note:* This figure summarizes the experimental design. After the baseline survey, respondents who express interest in the AI-powered news app are assigned to article environments that differ in political spin. For expositional convenience, the figure collapses the three politically neutral versions into a pooled *neutral* condition and shows the two political treatment arms as *liberal* and *conservative*. Respondents then use the app for a 10-day incentive period, complete the midline survey, and are observed for three additional 10-day post-incentive periods, of which the first is shown explicitly here before the endline survey. In the main analysis, the liberal and conservative assignments are recoded as *aligned* or *misaligned* political spin using respondents' partisan identification.

### 2.3.2 Recruitment Survey and Sample Selection

The recruitment survey serves two primary purposes. First, it allows us to identify and recruit a sample of individuals who express interest in accessing an AI-powered news application and own a compatible smartphone (iPhone or Android). Second, it collects baseline measures on demographics, media usage, and attitudinal characteristics.

The survey begins with a standard attention check and a question about the respondent's smartphone operating system. Respondents who fail the attention check or do not own a compatible device are screened out. The remaining participants answer questions on their news consumption habits and attitudes toward the media. These include retrospective measures of news outlet usage over the past 12 months, as well as items on news satisfaction, fatigue, interest, trust, and perceptions of current news reporting. The survey also elicits views on artificial intelligence, political attitudes (including measures of affective and issue polarization), mental health, and a rich set of demographic variables.

To elicit interest in the AI news app in a naturalistic way, we frame it as an opportunity to receive "free early access" prior to its public launch. Respondents are informed that the app provides frequent updates on trending stories based on Associated Press (AP) coverage, rewritten using AI technology "to deliver essential information quickly and clearly." We emphasize that the factual content remains unchanged, "guaranteeing the same level of trust and credibility you expect from seasoned journalists." Respondents explicitly consent to the collection of click data.

#### 2.3.3 App Download Survey

Several days after completing the initial recruitment survey, respondents who indicate interest in the app are invited to a follow-up survey related to app installation. In this subsequent survey, we verify respondents' continued interest in using the app. Those who confirm their interest are informed that they will receive a $2 bonus conditional on completing four tasks: downloading the app, logging in with provided credentials, finishing a short onboarding procedure, and enabling push notifications. Respondents who consent receive login credentials along with detailed installation instructions. Android users access the app directly via the Google Play Store, whereas iPhone users must first install TestFlight—a beta-testing platform—before downloading the app.

**Treatment assignment** Respondents are randomly assigned to one of five article environments: *Regular*, *Less Complex*, *Facts Only*, *Liberal*, or *Conservative*. They are not informed that different versions of the articles exist, nor are they told which version they are assigned. Any inference about the political slant of the content must therefore arise from their experience using the app. In the main analysis, as described in Section 3.1, we recode these assignments into aligned, misaligned, and pooled neutral conditions using respondents' partisan identification.

**Usage Incentives.** To ensure comparable levels of app-based news consumption across the aligned spin, misaligned spin, and neutral treatment conditions, we implement a series of usage incentives. First, the initial $2 download bonus is made conditional on enabling push notifications. Second, upon completing onboarding and activating notifications, respondents are informed of an additional incentive: they earn $1 for each day on which they read at least three articles in the app over a 10-day period. Third, eligibility for the initial notification bonus is verified over that same 10-day incentive period: respondents must keep notifications enabled throughout the period in order to receive the bonus. All usage-related incentives therefore end after day 10, so the post-incentive periods are not incentivized. Because we observe all in-app activity directly, compliance can be monitored without relying on self-reported data.

#### 2.3.4 Surveys

Midline and endline surveys are fielded 10 and 20 days after the app download survey, respectively, and the endline survey is virtually identical to the midline survey. We continue to observe app-based news consumption for an additional 20 days beyond endline, yielding three 10-day post-incentive periods in the usage data.

These surveys elicit measures of news satisfaction, self-reported news consumption outside the app, and time substitution, as well as perceptions of the app more broadly. They also collect data on affective and issue-based political polarization, factual knowledge, and mental health. We describe our primary outcome measures in detail below.

**News Satisfaction, Time Substitution, and App Perceptions** We assess news satisfaction by asking respondents to rate their overall satisfaction with the news they have read over the past week. To capture time substitution, we collect self-reported data on time spent consuming news from other sources, as well as changes in total news consumption relative to respondents' usual behavior. In addition, we offer a $2 incentive to upload a screenshot of phone screen time, providing a behavioral, non-self-reported measure of substitution patterns. To gauge perceptions of the app, we ask respondents to evaluate the trustworthiness, complexity, quality, entertainment value, political bias, factual accuracy, sentiment, and degree of opinion in the articles presented.

**Political Polarization** To measure affective polarization, we employ two instruments. First, we administer a standard feeling thermometer item, asking respondents how warmly or coldly they feel toward Republicans and Democrats. Second, we elicit willingness to share resources across party lines by asking respondents to divide a hypothetical $100 between a Democratic and a Republican voter from the most recent presidential election. To assess issue-based polarization, we collect respondents' policy views on a range of topics, including immigration, firearms, international trade, racial discrimination by police, and sexual harassment.

**Factual Knowledge** We measure factual knowledge using a 15-item multiple-choice quiz based on news stories shown during the incentive period. Each question offers five response options. Respondents earn a $0.20 bonus for each correct answer, conditional on answering in under 20 seconds to discourage web searching.

#### 2.3.5 Sample

**Setting** We recruited respondents through Prolific, an online survey platform widely used in the social sciences and known for its high data quality (Peer et al., 2022). Prolific is an online platform known for attracting younger, more educated, and tech-savvy respondents compared to the general U.S. population.

Prolific facilitates the payment of monetary bonuses and enables reliable re-contacting of participants for multi-wave studies. Data were collected in two waves, conducted in October/November 2024 and January/February 2025. The core study design was the same across waves, but the Facts Only arm was introduced only in the second wave. In addition, some survey questions in wave 2 were adjusted to reflect changes associated with the new presidential administration.

Among the 1,232 respondents who were assigned to our three treatments of interest, 76.2% participated in the midline survey and 71.7% participated in the endline survey (Figure B.5). There is no evidence of selective attrition by treatment status in both the midline and the endline survey (Table B.5). The downstream survey analyses are based on the subset of respondents who both participate in the relevant follow-up survey and can be classified as aligned or misaligned using partisan identification; this is why the estimation sample in the perception and polarization tables is smaller than the full randomized sample.

**Summary Statistics** Appendix Table B.3 reports summary statistics for the sample. Respondents have an average age of 37.3 years (range: 18–76), somewhat younger than the U.S. median of 38.9. Female respondents make up 57.1 percent of the sample, exceeding the national share of 50.8 percent. The sample is 68.3 percent White, broadly comparable to the U.S. average of 61.6 percent. Educational attainment is high: 87.0 percent report having at least some college education, relative to a national rate of 63 percent. The sample also leans more Democratic than the general population, with 43.0 percent identifying as Democrats (vs. 30–35 percent nationally), 33.2 percent as Independents, and 23.9 percent as Republicans. In terms of media consumption, 90.3 percent primarily access news online—substantially higher than the national figure of 58 percent—and most respondents consume news at least somewhat frequently, and use an average of 1.8 news apps. Taken together, the Prolific sample is not nationally representative: it skews younger, more female, more educated, and more digitally engaged than the general U.S. population. Appendix Table B.4 provides evidence of covariate balance across the different treatment groups.

## 2.4 The Effects of Political Spin

In this section, we study the effects of political spin on news consumption, perceptions of the news app, factual knowledge, and political polarization.

### 2.4.1 News Consumption

**Empirical Specification** To study the effects of political spin on news consumption, we estimate the following equation:

$$y_{it} = \alpha + \beta \text{Aligned}_i + \gamma \text{Misaligned}_i + \delta X_i + \varepsilon_{it} \tag{2.1}$$

$y_{it}$ is respondent $i$'s news consumption in period $t$. We define one period to be a 10-day horizon to match the duration of the incentive period. We consider news consumption data up until 30 days after the incentive period, which gives us a total of four periods: the 10-day incentive period and three subsequent 10-day post-incentive periods. In the discussion below, we first summarize treatment differences separately for the incentive period and for the pooled post-incentive period, and then turn to the raw period-by-period data. To classify participants as either identifying with Democrats or Republicans, we rely on a two-step question. First, respondents are asked which political party (Democrats, Republicans, Independents) they identify with. Those who identify as Independents are then asked whether they lean Democrat or Republican.[3] $\text{Aligned}_i$ takes value one for respondents who identify with the Democratic party and are assigned the liberal rewrites, and respondents who identify with the Republican party and are assigned

[3] In Appendix Section B.13, we assess the robustness of our results using an alternative measure of political alignment. Specifically, we rely on respondents' answers to the question, "Are you liberal or conservative?" with five response options: *Very conservative*, *Conservative*, *Neither conservative nor liberal*, *Liberal*, and *Very liberal*. The main findings are very similar when using this measure to define alignment.

the conservative rewrites, and zero otherwise. Respondents who do not identify with either party and do not report a partisan lean are excluded from these aligned-versus-misaligned specifications. Conversely, $\text{Misaligned}_i$ takes value one for respondents who identify with the Democratic party but are assigned the conservative rewrites, and respondents who identify with the Republican party but are assigned the liberal rewrites, and zero otherwise. As discussed in Section 2.2.2, we use respondents who receive politically neutral versions as the control group in these analyses.[4] $X_i$ is a vector of control variables (age, gender, race, income, some college, full-time employment, primarily consuming news online, having a paid news subscription, consuming news at least somewhat frequently, number of news apps, primarily reading news on the phone, trust in news and trust in AI, and partisan identification). $\varepsilon_{it}$ captures an individual-period error term. Standard errors are clustered at the respondent level.

**Incentive Period** The top panel of Figure 2.4 (and Table B.8) shows differences in news consumption during the incentive period, comparing respondents in the aligned and misaligned conditions to respondents in the neutral conditions. We focus our analysis of news consumption on two main outcomes, which we standardize using the mean and standard deviation of the neutral rewrite groups: the share of days with any activity in the app and winsorized daily time spent in the app.

The figure shows muted and statistically insignificant effects on both days with app use and time spent. The effect sizes for days and time spent are relatively close to zero for respondents in the aligned spin condition. The point estimates indicate that respondents log into the app on 0.08 SD fewer days in the aligned spin condition and on 0.07 SD more days in the misaligned condition, but neither coefficient is statistically significant ($p = 0.253$ and $p = 0.370$, respectively). For time spent, the coefficients are very close to zero. Taken together, these results underscore that the incentives ensured similar levels of overall consumption across treatment conditions. Appendix Table B.7 shows a related but more qualified pattern for exposure to the focal app categories (*Politics*, *Economy*, *U.S.*, and *World*): focal-category time is similar in levels across groups, but both treated groups devote slightly smaller shares of reading time to these categories than the pooled neutral group, and the aligned group also opens fewer focal-category articles on average. In short, the incentive design equalizes overall app use more clearly than it equalizes the exact dose of politically spun content. We therefore interpret the midline comparisons with the qualification that some substitution toward neutral categories may remain during the incentive period.

**Post-Incentive Period** The bottom panel of Figure 2.4 shows that news consumption is significantly lower when respondents receive misaligned spin. In particular, the days with app use and the daily time are each 0.13 SD lower in the misaligned condition than in the neutral condition ($p < 0.01$), respectively. For time spent, this corresponds to about 0.16 fewer minutes per day (roughly 9.5 seconds), or about 1.6 fewer minutes over a 10-day post-incentive period. Appendix Figure B.9 shows that the effects of alignment get stronger over the course of the post-incentive period. These findings are consistent with models that posit that respondents

[4]As we show in Appendix Section B.12, respondents in these different politically neutral groups hold similar beliefs and exhibit similar consumption patterns. This provides further justification for our approach of pooling these conditions.

have a distaste for politically misaligned news (Gentzkow and Shapiro, 2006; Mullainathan and Shleifer, 2005).

At the same time, news consumption is not significantly higher for respondents receiving aligned spin compared to the neutral control group. The point estimates are slightly negative in the aligned condition: respondents have 0.05 SD fewer days with app use ($p = 0.305$) and spend 0.02 SD less time in the app ($p = 0.625$). The point estimate for time corresponds to only about 0.03 fewer minutes per day, underscoring that the aligned effect is economically small and statistically indistinguishable from zero. The absence of any positive consumption response to aligned spin stands in contrast to simple symmetric demand-side predictions. Relative to the neutral control, aligned spin does not generate a detectable increase in engagement.

**Raw data** Figure B.6 presents the raw usage data by treatment group.[5] The figure shows no meaningful consumption differences across treatment groups during the incentive period, consistent with the role of the usage incentives in equalizing exposure.

By contrast, substantial differences emerge once incentives are removed. Usage declines sharply across all groups in the post-incentive period, but remains persistently lowest for respondents exposed to politically misaligned spin. The control group exhibits the highest raw usage in the first post-incentive period, followed by the aligned group, with the misaligned group clearly lowest. The same ranking broadly persists over the subsequent 20 days as overall usage decays. These raw patterns mirror the regression results and underscore the asymmetric effect of political spin on sustained news consumption.

### 2.4.2 Perceptions and News Satisfaction

**Perceptions** To understand why exposure to politically misaligned spin reduces news consumption, while aligned spin fails to increase it, we turn to rich perception data collected in the midline survey.[6]

To characterize effects on perceptions, we re-estimate equation 2.1, except that the outcome variables are now beliefs measured in the midline survey. Figure 2.5 and Table B.9 present treatment effects on perceptions of the news content, with all outcomes standardized to z-scores. Respondents assigned to politically aligned spin report limited differences relative to the neutral condition. They clearly perceive the content as more politically aligned with their own views—by 0.42 SD—but do not rate it as more accurate (0 SD), trustworthy (−0.02 SD), or satisfying (−0.01 SD).[7] The one notable positive perception effect is on perceived quality, which increases by 0.17 SD and is statistically significant at the 5% level. Respondents also view the aligned content as modestly more biased (0.14 SD), suggesting that they recognize the ideological slant but do not interpret it as undermining quality.

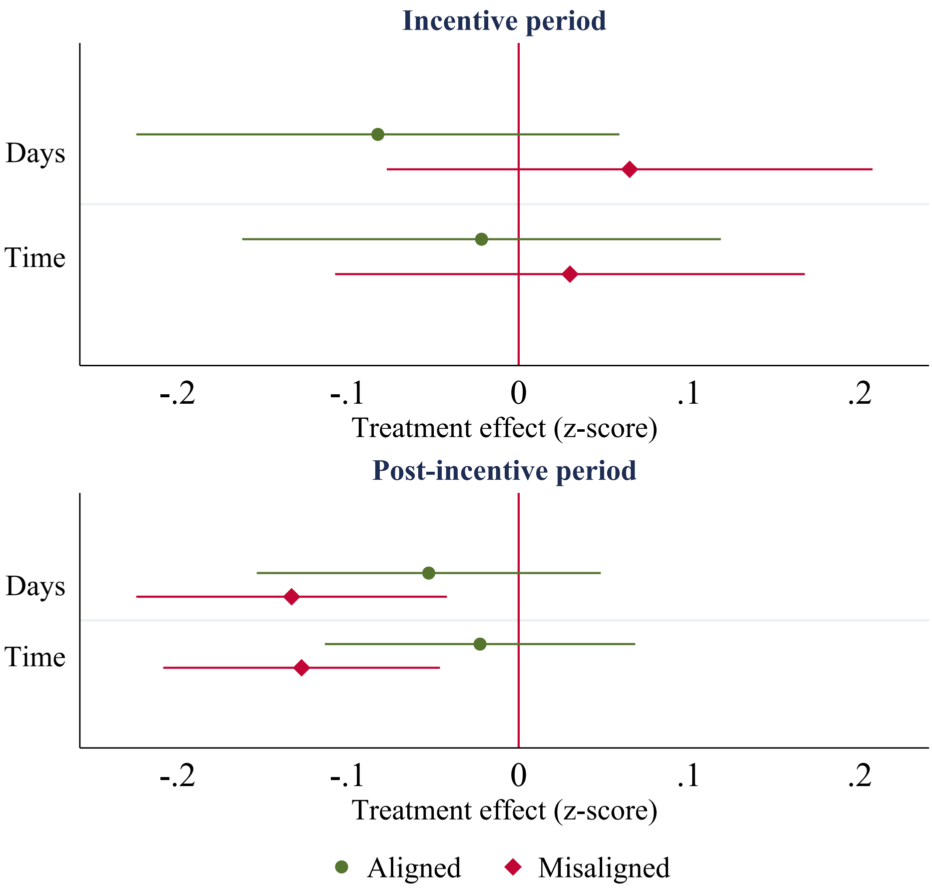
By contrast, exposure to misaligned spin leads to substantially more negative perceptions across all evaluated dimensions. Respondents rate the content as less accurate (−0.38 SD), less trustworthy (−0.39 SD), and lower in overall quality (−0.19 SD); the first two effects are

[5]Figures B.7 and B.8 show analogous patterns for the share of days using the app and the number of articles read.

[6]In Appendix Figure B.19, we show that our results on perceptions are fairly similar in the endline survey.

[7]We define political alignment as a variable taking larger values when the perceived political bias of the outlet aligns with respondents' political orientation, and smaller values when it is misaligned. Specifically, we use a 5-point scale from "very right-wing biased" to "very left-wing biased" and reverse-code the variable for Republicans.

Figure 2.4: Effects on News Consumption




*Note:* This figure shows the effect of aligned and misaligned versions on news consumption in the incentive period (top panel) and post-incentive period (bottom panel). News consumption is measured as the share of days using the app and winsorized daily time spent using the app. Both outcomes are standardized. Error bars represent 95% confidence intervals.

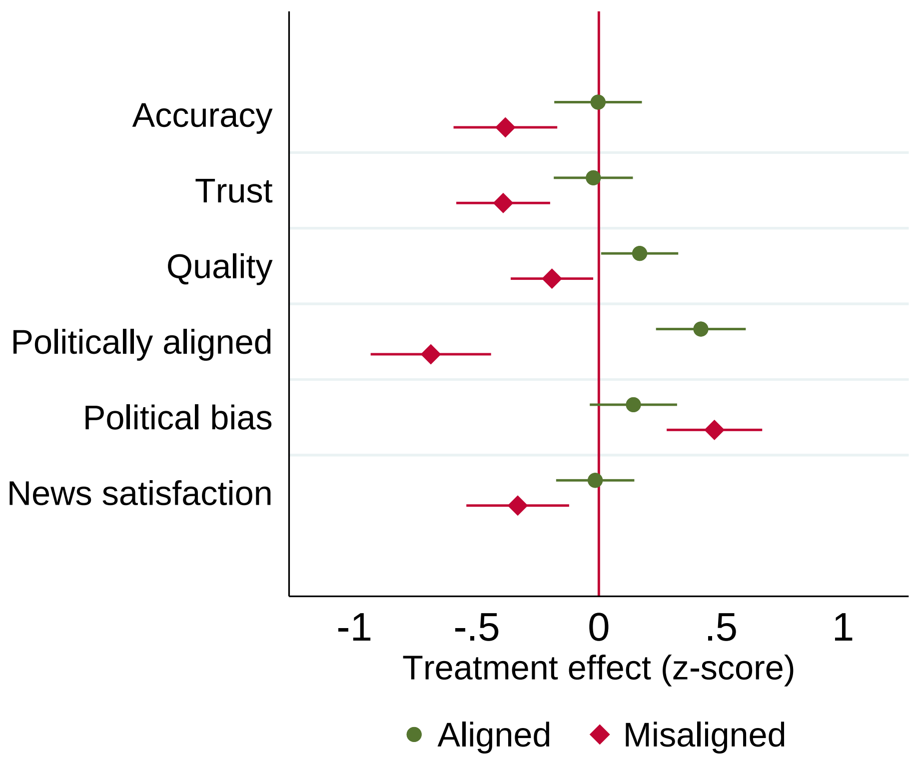
significant at the 1% level and the quality effect is significant at the 5% level. Satisfaction also declines by 0.33 SD ($p < 0.01$). Misaligned spin is perceived as both substantially less politically aligned with the respondent's own views (−0.69 SD) and significantly more biased overall (0.47 SD). These results are consistent with Bayesian models of source credibility under uncertainty (Gentzkow and Shapiro, 2006), and help explain why misaligned spin reduces news consumption while aligned spin fails to increase it: misalignment undermines perceived credibility and satisfaction, while alignment, despite being recognized, does not improve perceived accuracy, trustworthiness, or satisfaction relative to a neutral baseline, even though it does modestly raise perceived quality.[8] This null aligned effect should be interpreted relative to the neutral control, which is close to AP on the political dimension and yields similar results across its component versions (Appendix Figure B.4 and Appendix Section B.12). Appendix Figure B.19 shows that these patterns are broadly similar in the endline survey: misaligned spin continues to reduce perceived accuracy, trust, and news satisfaction, while aligned spin still does not

[8]Our findings are broadly consistent with evidence showing that politically neutral scientists are perceived to be more credible than scientists of aligned or misaligned leaning (Alabrese et al., 2026).

produce meaningful gains in satisfaction relative to the neutral condition. Consistent with this interpretation, Appendix Table B.10 shows that respondents who report higher perceived accuracy and news satisfaction at midline also consume more news in the post-incentive period, although these correlations are descriptive and should not be interpreted causally.

**Spillovers to Other Perceptions** We next examine whether respondents exposed to aligned or misaligned spin update their beliefs about other characteristics of the app's content. Appendix Figure B.18 shows little evidence of spillovers to other perceptions. Perceived complexity, entertainment, and sentiment are similar across groups, while trust in AI rises modestly in the aligned condition (0.16 SD) but is unchanged in the misaligned condition. These patterns cannot explain the consumption results; if anything, the increase in trust in AI under aligned spin would predict higher, not lower, consumption relative to the neutral control.

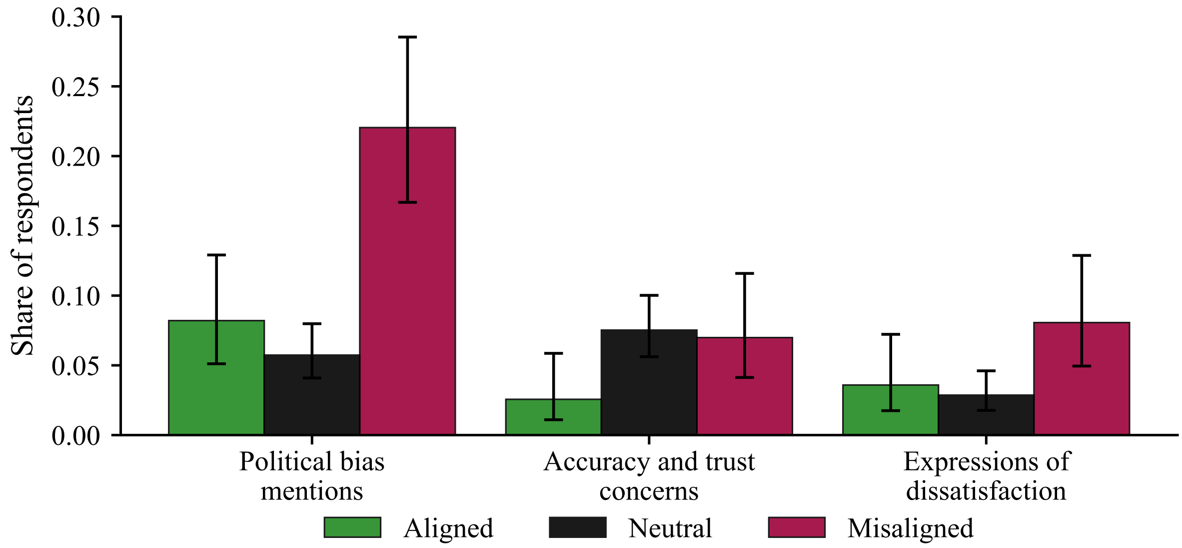
Figure 2.5: Effects on Perceptions and News Satisfaction




*Note:* This figure shows the effect of aligned and misaligned versions on perceptions of the app and news satisfaction. *Accuracy*, *Trust*, and *Quality* measure respondent ratings of the articles in the app. Political ratings are based on respondent ratings of articles on a 5-point scale from "very right-wing biased" to "very left-wing biased". *Politically aligned* is defined as proximity to "very right-wing biased" for Republicans and "very left-wing biased" for Democrats. *Political bias* is an indicator for rating it as biased in any direction (as compared to "not biased"). *News satisfaction* is respondents' reported satisfaction with the news they have been reading over the past week. All outcomes are standardized. Error bars represent 95% confidence intervals.

**User Experience Data** In our surveys, respondents receive an open-ended question in which they were asked to describe their experience with the app. We analyze this data to shed additional light on respondents' perceptions of the app on top of the previously discussed quantitative measures. The open-ended responses are informative because they shed light on the spontaneous associations respondents report when reflecting on the app (Haaland et al., 2025). To analyze these responses, we use an LLM-assisted coding procedure that classifies the free-text answers into a set of categories; Appendix Table B.14 reports the coding scheme, and Appendix Figure B.27 shows the full set of coded categories across treatment groups. Figure 2.6 focuses on the three categories that are most informative for the main mechanism. The clearest pattern is that respondents exposed to politically misaligned spin are more likely to describe the content as politically biased: 22% of respondents in the misaligned group mention political bias, compared with 6% in the neutral group and 8% in the aligned group. Misaligned respondents are also more likely to express dissatisfaction with the app, whereas aligned respondents look broadly similar to the neutral group on this dimension. What does not rise one-for-one in the open-ended data is spontaneous discussion of accuracy or trust. We therefore read Figure 2.6 as complementing, rather than mechanically mirroring, the structured survey responses: the strongest spontaneous reaction to misaligned spin is to notice ideological slant and report dissatisfaction, while the more specific trust and accuracy downgrades are primarily visible in the direct perception questions.

Figure 2.6: Analysis of Open-Ended Responses




*Note:* This figure shows the share of respondents in each treatment whose open-ended response was coded as containing the indicated theme. The three categories shown are only a subset and all categories are shown in Appendix Figure B.27. Responses can receive multiple codes, so percentages need not sum to 100 across themes. Error bars represent 95% confidence intervals.

### 2.4.3 Knowledge Acquisition

To assess how political spin influences knowledge acquisition, we analyze treatment effects on respondents' performance in an incentivized factual quiz administered during the midline survey.[9] Previous research suggests that exposure to ideologically congruent or incongruent

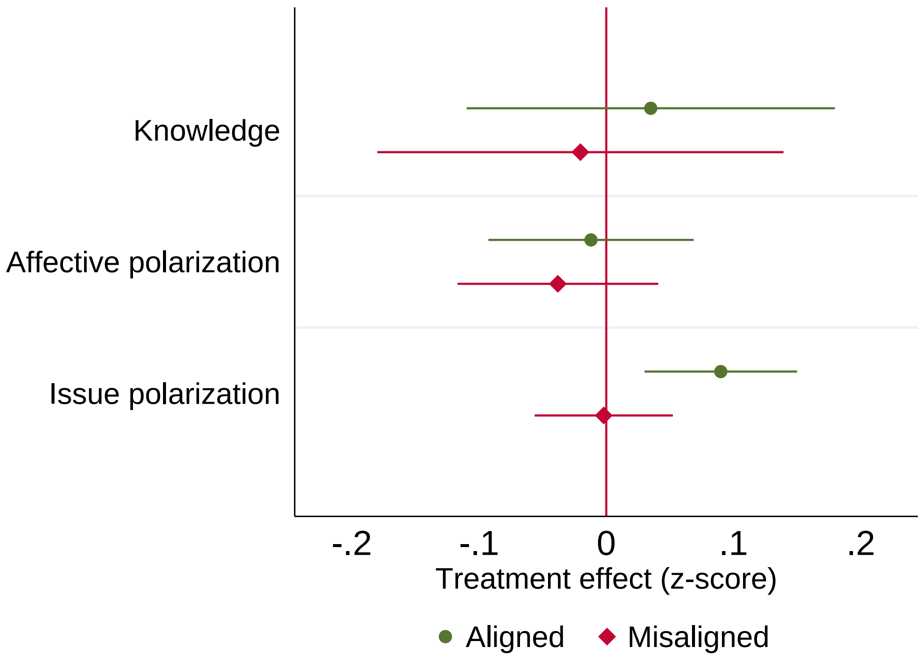
[9] Appendix Figure B.20 shows similar treatment effects on knowledge acquisition in the endline survey.

news can alter knowledge acquisition, either by reinforcing existing beliefs or prompting critical scrutiny of the content (e.g., Druckman et al., 2013; Nyhan and Reifler, 2010). In our setting, however, large effects on factual knowledge are not especially likely: the treatment changes framing, tone, and rhetorical emphasis while holding the underlying news events constant. Consistent with this design, we find only muted effects on knowledge. Respondents assigned to aligned spin score 0.04 standard deviations higher than those in the neutral group, while those exposed to misaligned spin score 0.02 standard deviations lower; neither difference is statistically significant (Figure 2.7 and Table B.11). These findings indicate that, despite substantial effects on perceptions and consumption, political spin has limited influence on objective knowledge acquisition in our context. Thus, the absence of detectable knowledge effects is consistent with a treatment that changes how news is presented more than what information respondents are exposed to. Appendix Figure B.20 shows that these downstream patterns are also broadly similar at endline: we continue to find no clear effects on knowledge or affective polarization, while the positive effect of aligned spin on issue polarization remains visible but should be interpreted cautiously in light of the broader pattern of mostly null downstream results.

### 2.4.4 Polarization

Figure 2.7 and Appendix Table B.11 report treatment effects on political polarization as measured in the midline survey right after the incentive period, after a period in which overall app use had been partially equalized by incentives. Exposure to politically aligned spin results in a modest increase in issue polarization (0.09 SD). Appendix Table B.12 shows that this is the only downstream outcome that remains statistically significant after adjusting across the three headline downstream outcomes in the paper—knowledge, affective polarization, and issue polarization—with a Holm-adjusted $p = 0.010$ and a Benjamini-Hochberg $q = 0.010$ for the aligned effect. The corresponding misaligned estimate is close to zero and not statistically distinguishable from zero. Appendix Table B.13 then decomposes the issue-polarization result into the five issue components that enter the index and reports component-level Holm-adjusted p-values and Benjamini-Hochberg q-values. The strongest aligned effect appears for international trade (0.24 SD, Holm-adjusted $p = 0.028$, BH $q = 0.028$). Sexual-harassment attitudes also move in the polarizing direction at the raw 5% level, but that component does not survive multiplicity adjustment. The remaining aligned component estimates are smaller and statistically indistinguishable from zero, and none of the misaligned component estimates survive adjustment. We therefore interpret the composite result as evidence of a modest increase in issue polarization, but not as a uniform shift across all five policy domains. By contrast, neither treatment meaningfully alters affective polarization, indicating that changes in perceived policy distance do not translate into heightened partisan animus. Given that we measure both affective polarization and issue polarization already at baseline, we have relatively high levels of statistical power. This allows us to rule out relatively small effect sizes for both issue and affective polarization.

Figure 2.7: Effects on Knowledge and Polarization




*Note:* This figure shows the effect of aligned and misaligned versions on knowledge and polarization. *Knowledge* refers to the score in a quiz of 15 questions based on recent news stories. *Affective polarization* is an index combining feeling warmly toward Republicans and Democrats and hypothetical division of $100 between a Democrat and Republican voter. *Issue polarization* is an index combining policy views on immigration, firearms, international trade, racial discrimination by police, and sexual harassment. All outcomes are standardized. Error bars represent 95% confidence intervals.

### 2.4.5 Heterogeneity by Ideology

**Political Heterogeneity** Which respondents drive the asymmetric average effects documented above? Heterogeneity analyses point to a common pattern: the main response margin is not strong positive demand for aligned spin, but rather aversion to misaligned spin among respondents with stronger political commitments.

Figures B.10 and B.12 first show treatment effects separately for respondents leaning Democrat and those leaning Republican. The broad asymmetry is visible for both partisan groups. Misaligned spin reduces perceived accuracy, trustworthiness, quality, and news satisfaction across parties, whereas aligned spin generates only limited improvements beyond perceived political alignment. On the consumption side, the most notable post-incentive response is that Democrats exposed to misaligned spin reduce app usage significantly, while aligned spin again does not generate a corresponding increase. Among Republicans, the point estimates during the incentive period suggest some responsiveness to alignment, but these differences do not translate into a comparably clear post-incentive increase in engagement. Therefore, the party splits are consistent with the broader pattern that misaligned spin drives the main behavioral downside, whereas aligned spin produces at most modest benefits.

**Strength of Ideology.** To study more directly how political commitment shapes these responses, we split respondents by whether they identify as liberal or conservative versus "neither conservative nor liberal." This distinction yields the clearest mechanism evidence. Figure B.11 shows that during the incentive period, both ideological and non-ideological respondents consume similar amounts of news regardless of alignment. Once incentives are removed, however, misaligned ideological respondents substantially reduce their consumption, with significant negative effects on both usage days and daily time spent (around –0.18 to –0.19 SD), while aligned ideological respondents exhibit no increase. Among non-ideological respondents, treatment effects are close to zero for both aligned and misaligned spin.

Figure B.13 shows an analogous pattern for perceptions and news satisfaction. Among ideological respondents, exposure to misaligned spin sharply reduces perceived accuracy, trust, quality, and news satisfaction (–0.3 to –0.5 SD). By contrast, aligned spin mainly increases perceived political alignment without delivering meaningful gains in quality or satisfaction. For non-ideological respondents, responses are muted across the board, with smaller and statistically insignificant effects on both consumption and perceptions. This contrast suggests that the aggregate asymmetry is concentrated among respondents with stronger political commitments.

Appendix Figures B.14 and B.15 further reinforce this interpretation by unpacking heterogeneity by partisan strength. The consumption results indicate that the post-incentive reduction in app use following misaligned spin is larger among strong partisans than among non-strong partisans. For perceptions and satisfaction, the point estimates are generally also more negative for strong partisans, although these differences are imprecisely estimated and should be interpreted with caution. By contrast, Appendix Figures B.16 and B.17 show little systematic heterogeneity in knowledge and polarization outcomes by party or ideology, consistent with the broader pattern of limited downstream effects.

In sum, the heterogeneity evidence points to a common mechanism. The asymmetric response to political spin is not driven by a broad demand for politically aligned framing. Instead, the dominant force in the data is aversion to misaligned spin among respondents with stronger political commitments, for whom counter-attitudinal framing is especially costly in terms of perceived credibility and satisfaction. Among weaker or non-ideological respondents, treatment effects are substantially smaller and often close to zero. This pattern helps explain why the average effect of misaligned spin is pronounced while the average effect of aligned spin remains muted: the key response margin is disengagement from counter-attitudinal framing rather than increased demand for aligned spin.

### 2.5 External Validity

The external validity of our findings depends on three main features of the setting. First, the study sample is drawn from Prolific and further selected on respondents being willing to download and use an AI-powered news app. Relative to the broader U.S. population, this likely yields a sample that is younger, more educated, more digitally engaged, and more open to AI-mediated news consumption. Appendix Table B.6 shows that reported willingness to download the app is high overall in our experimental sample and only weakly related to observable baseline characteristics. In addition, the study takes place in the United States during a highly salient political period spanning late 2024 and early 2025. Effects could therefore differ in less polarized periods, in lower-salience news environments, or in countries with different party systems and media institutions.

Second, our setting deliberately abstracts from several features that are central to real-world partisan media environments. Respondents know they are using an AI-powered news app, but they are not told that different political rewrites of the same stories exist across users. Any updating about slant, credibility, or source quality must therefore arise from the reading experience rather than from explicit treatment labels or prior beliefs about outlet identity. This is a strength for identifying how framing shapes perceptions under uncertainty, but it also means that our design strips away outlet brands, stable reputations, parasocial attachment, and the social meaning attached to consuming a particular source. Moreover, the intervention takes place inside our own app using articles based on the Associated Press, a source widely regarded as highly credible and comparatively nonpartisan. Our estimates therefore speak most directly to the marginal effect of political spin in a high-credibility, low-brand-signal environment, and they may understate the effects of aligned framing in settings where it is layered onto a lower-credibility or already slanted baseline.

Third, our estimates identify the short- to medium-run effect of framing alone, not the full effect of sustained partisan media exposure. By design, we vary political framing while holding the underlying news event and coverage fixed. This is a key strength for causal identification, but it also implies that our estimates should not be interpreted as the total effect of partisan media in practice. Real-world partisan media environments typically combine framing with differences in story selection, repetition, agenda setting, and source reputation, and these margins may reinforce one another. In addition, although we observe respondents during a 10-day incentive period and for three subsequent 10-day post-incentive periods, this horizon may still miss longer-run dynamics such as habit formation, cumulative learning, or the gradual build-up of polarization. Our evidence is therefore most informative about the short- to medium-run effects of framing holding coverage fixed, rather than the long-run effects of persistent exposure to full partisan media ecosystems.

### 2.6 Conclusion

This paper studies the causal effects of political spin in news reporting while holding the underlying news event fixed. Using a field experiment embedded in an AI-powered news app, we compare respondents exposed to aligned, misaligned, and politically neutral versions of the same underlying stories. This design allows us to isolate the effect of ideological framing from differences in story selection or coverage.

We uncover a clear asymmetry in how audiences respond to political spin. Relative to the neutral control, exposure to misaligned spin lowers post-incentive news consumption and substantially worsens perceptions of accuracy, trustworthiness, quality, and satisfaction. By contrast, aligned spin does not increase news consumption or satisfaction relative to that benchmark, even though respondents clearly recognize it as more politically aligned. The core pattern is therefore punishment without reward: political spin appears to generate substantial downside risk when it is misaligned, but little offsetting benefit when it is aligned. The subgroup evidence suggests that this asymmetry is driven primarily by aversion to misaligned spin among respondents with stronger political commitments rather than by strong positive demand for aligned spin. Turning to downstream consequences, we find limited effects on factual knowledge and affective polarization. At the same time, aligned spin leads to a modest increase in the composite issue-polarization measure that remains significant after adjusting across the paper's three headline downstream outcomes, with the component evidence pointing most clearly to trade-related attitudes rather than to a uniform shift across every issue. The overall pattern therefore points to meaningful behavioral and attitudinal effects of political framing, but in a form that is narrower and more asymmetric than a simple model of demand for like-minded news would suggest.

These findings have implications for how we think about the market for news. If politically aligned spin does not raise either engagement or satisfaction, then simple demand-side explanations based on spin alone appear incomplete. At the same time, our findings point toward a larger role for other forces that shape partisan media environments, including story selection, outlet identity, and supply-side incentives. More broadly, the paper shows that framing alone, even without changing which events are covered, can shape news consumption, perceived credibility, and issue polarization.

Chapter 3

# Measuring the Structure of Household Macroeconomic Expectations

*by Fabian Roeben*

**Abstract:** Households differ not only in the levels of their macroeconomic expectations but also in how they relate these expectations. Using 6,550 respondents observed for 12 monthly waves in the NY Fed Survey of Consumer Expectations, I measure within-household comovement among four year-ahead expectations: inflation, unemployment, interest rates, and stock prices. Expectations comove far more strongly than noisy answering alone can explain: an index of how tightly expectations move together is stable enough to rank households, where inflation is the least connected factor. The measure remains visible after accounting for expectation levels, demographics, and response style, and its sign pattern differs from an optimism-pessimism index: households reporting a higher chance that unemployment will rise also report a higher chance that stock prices will rise in the same months. Comparable within-household comovement appears in the ECB Consumer Expectations Survey under different questions, response scales, and macroeconomic conditions. Furthermore, direct prediction of expected spending levels is limited, but the measure changes how inflation expectations should be interpreted: the inflation–spending slope is about one quarter flatter among households whose expectations move most tightly together, consistent with linked adverse expectations offsetting part of the positive inflation–spending association. Therefore, expectations surveys contain a measurable and behaviorally relevant layer beyond expectation levels that is not typically used.

*Keywords:* Macroeconomic expectations, Within-household comovement, Subjective expectations, Survey of Consumer Expectations, Inflation expectations

*JEL codes:* C38, C83, D83, D84, E21, E31

## 3.1 Introduction

A household whose inflation and unemployment expectations are both high on average differs from a household whose inflation and unemployment expectations move together over time: the first fact is about levels, the second about structure. A large literature studies what households expect inflation to be, whether unemployment expectations are accurate, or how expected inflation maps into spending, typically one expectation at a time. That approach is useful when the goal is a single expectation or a single behavioral margin. It leaves open a different question: do household macro expectations move together in household-specific ways, and if so, what does that structure reveal?

This paper studies that question in repeated survey data. I refer to belief-system structure as the way a respondent's own macro expectations move relative to one another over time, and I measure one observable component of it: how strongly those expectations comove systematically within person across survey waves. One household may move several macro expectations together; in another the same expectations may move largely independently or in opposite directions; the strength of that linkage is itself a household characteristic.

Policy makers and researchers often interpret one reported expectation without observing the other expectations reported by the same household. Two households can report the same average inflation expectation and still connect that expectation to different movements in unemployment, interest-rate, or stock-market expectations. If so, expectation levels alone miss part of the information contained in expectations surveys. For central banks communicating about inflation, unemployment, or financial conditions, the same message may reach households with very different patterns of linked expectations, and a single reported expectation can carry different meaning depending on what else moves with it. A panel measure of comovement also gives a way to study linked expectations without directly eliciting respondents' subjective causal models or imposing a researcher-specified macro model on their answers. Repeated expectations surveys may then carry information not only about each expectation separately but also about the links among expectations.

The paper asks three questions. Do macro expectations comove within household, and is that comovement stable enough to rank or classify households? Does the same structure appear beyond one survey and predict how households organize their expectations later? And does it change how individual expectation levels map into behavior?

The empirical design starts from a common panel. I use the NY Fed Survey of Consumer Expectations (SCE) and construct respondent panels with 12 complete monthly survey waves, giving roughly 6,500 households observed over the same panel length. For each respondent, I compute the six within-household correlations among four one-year macro expectations: the average inflation expectation implied by the respondent's reported probability distribution over inflation outcomes (the density-implied mean), the probability unemployment will be higher, the probability saving-account rates will be higher, and the probability stock prices will be higher. These six correlations are the starting point. I then summarize them using simple continuous indices and ask which summaries survive basic validation.

The main result is that household panels contain systematic heterogeneity in how macro expectations move together. In the SCE, the strongest signal is a continuous ranking of households by how tightly three probability expectations move together: unemployment risk, saving-account rates, and stock prices. At one end of the ranking, the three probabilities move closely together; at the other end, they move independently or in opposite directions. Figure 3.1 shows example respondents from both ends. This joint movement does not look like simple optimism or pessimism. In months when a household reports a higher chance that unemployment will rise, it also tends to report a higher chance that stock prices will rise, the opposite of a simple good-times-versus-bad-times factor. Instead, some respondents appear to move several assessments of macro change together: when they report that one macro outcome is more likely, they often report that others are too, even when the directions do not fit a single optimistic or pessimistic story.

Inflation plays a different role in this structure. Inflation-link comovement is positive on average, with a mean of 0.033, but much less stable than the probability block when respondent histories are split. The instability is itself informative: under every inflation elicitation the survey offers—including when inflation is re-expressed as a probability on the same 0–100 scale as the other three inputs—the expectation most central to policy communication is the one least reliably connected to the rest of the household's macro beliefs.

Among the mechanisms that could generate joint probability movements, perceived uncertainty provides the clearest observable clue in the SCE. The survey asks respondents for a probability distribution over future inflation, so the width of that distribution measures reported inflation uncertainty. Within respondent, months with wider reported inflation densities are months in which several probability answers rise together, and the link strengthens across the comovement ranking. This fits the idea that some households sometimes see the macro environment as broadly unsettled: several outcomes feel more likely to change at once. Because the density IQR captures inflation uncertainty rather than overall macro uncertainty, I use it as mechanism evidence rather than as a decomposition of the index.

Validation turns these within-household correlations into a usable measurement object. With 12 monthly answers per respondent, the relevant question is whether the cross-household ranking is large relative to noise, stable across independent pieces of the panel, and distinct from expectation levels, demographics, and response styles. The average within-household correlation among the three probabilities is 0.222, far above shuffled-time benchmarks, and the cross-household ranking persists when respondent histories are split in half. The implied 12-wave reliability is 0.565, informative for cross-sectional comparisons even though individual scores remain noisy. Expectation levels, demographics, and response styles explain under a third of the observed variation, and the same respondents' comovement in non-macro probability questions is nearly uncorrelated with the macro index. Principal components recover the same probability block, while discrete household types do not replicate across independent halves of the same respondent histories. The SCE supports a continuous household-level ranking rather than a small set of durable household types. The cross-survey validation asks whether linked expectations appear outside the SCE. The ECB Consumer Expectations Survey (CES) is informative precisely because it is not a copy of the SCE: while broadly similar, it asks different questions, uses different response scales, covers several euro-area countries, and spans a different macroeconomic environment. Within-person comovement of adverse macro expectations also lies far above noise benchmarks in the CES and survives country-wave standardization, winsorization, and rank-correlation checks. The comparison shows that linked macro expectations are not a peculiarity of one U.S. questionnaire. The CES result is weaker at the individual level: respondent rankings are less reliable there, so I interpret the CES evidence as distributional portability rather than a second validated household-level ranking.

Within the SCE, the measure also predicts later expectation dynamics. Comovement measured in the first half of each respondent's panel predicts later joint probability movements and same-direction probability revisions, conditional on first-half expectation levels, demographics, and response style. This timing-separated exercise keeps the validation target close to the measurement object: it asks whether households that organize probability expectations together early in the panel continue to do so later.

Finally, the structure changes how inflation expectations map into behavior. The framework developed in Section 3.3 implies that the inflation–spending slope should depend on the expectation structure around inflation, and it does. Moving from the 25th to the 75th percentile of comovement flattens the implied slope by about a quarter, from 1.78 to 1.31 percentage points of expected spending growth per standard deviation of inflation expectations. This heterogeneity is concentrated in respondent windows from 2022 onward, which is where the mechanism should be visible—inflation was salient and the baseline inflation–spending slope was steepest—while the comovement measure itself appears in all timing windows. Comovement does not predict the level of expected spending once levels and response styles are controlled; its behavioral content is in the slope, not the intercept.

This paper makes three contributions. First, it defines and measures a household-level belief-system structure: the within-person comovement of repeated macroeconomic expectations. This shifts the object of analysis from isolated expectation levels to the links among expectations, using a transparent measure that can be constructed from standard public panel surveys. Second, it validates that measurement object. The SCE evidence shows that probability-expectation comovement is well above noise benchmarks, stable enough to rank households, and remains visible after accounting for average expectations, demographics, and response style; by contrast, inflation-link comovement and discrete type labels fall short of the same validation standard. Third, it maps where the measure matters. A related comovement pattern appears in the ECB CES under different questions and conditions; early SCE comovement predicts how households organize and revise expectations later; and comovement flattens the inflation–spending slope by about a quarter in the high-inflation period, even as it leaves the level of spending unexplained. The behavioral content is specific: the measure changes how a canonical inflation coefficient should be read, without claiming a direct effect on spending.

The rest of the paper proceeds as follows. Section 3.2 places the paper in the literature. Section 3.3 explains why joint movements in the four expectations are economically meaningful without imposing a single textbook sign pattern. Section 3.4 describes the data and measurement design. Section 3.5 presents the SCE measurement results. Section 3.6 validates the SCE measure. Section 3.7 studies the ECB CES comparison. Section 3.8 asks what the measure predicts. Section 3.9 concludes.

## 3.2 Related Literature

**Subjective macro models and linked expectations.** This paper is closest to work showing that households connect macroeconomic outcomes in heterogeneous and sometimes nonstandard ways. Evidence on subjective macro models, theory consistency, narratives, and lay economic reasoning shows that respondents may link inflation, unemployment, policy, output, and personal outcomes differently from textbook representations (Andre et al., 2022, 2026; Binetti et al., 2024; Carvalho and Nechio, 2014; Dräger and Lamla, 2024; Dräger et al., 2016; Kuchler and Zafar, 2019; Leiser and Aroch, 2009; Leiser and Krill, 2017; Roth and Wohlfart, 2020). My contribution is complementary. Existing studies often elicit causal links directly; I ask whether standard public panels that repeatedly ask about several outcomes contain stable empirical evidence that a respondent's reported expectations are linked. I use belief-system structure only in this narrow empirical sense: interdependence among reported expectations, not a directed subjective model.

**Expectations measurement and disagreement.** Household expectations surveys are central because they measure beliefs directly rather than inferring them from choices or aggregate outcomes (Armantier et al., 2017; D'Acunto and Weber, 2024; Weber et al., 2022). They also show persistent and systematic disagreement across households (Mankiw et al., 2003). I add a related form of disagreement: households differ not only in expectation levels but also in how their own expectations move together over time. Because survey answers are noisy and shaped by elicitation format, attention, uncertainty, and focal or rounded response behavior (Bruine de Bruin et al., 2011, 2012; Dominitz and Manski, 1997; Engelberg et al., 2009; Hurd, 2009; Manski, 2004, 2018; Manski and Molinari, 2010), reliability, tail handling, focal responses, and overlap with levels and demographics are core parts of the measurement design.

**Heterogeneity, updating, and dimension reduction.** Research on information frictions, learning, experiences, socioeconomic status, and endogenous information acquisition shows why respondents may differ in how expectations move when they process news or form beliefs (Andrade et al., 2023; Armantier et al., 2016; Carroll, 2003; Coibion and Gorodnichenko, 2012; Coibion et al., 2022; Das et al., 2020; Fuster et al., 2022; Kim and Binder, 2023; Link et al., 2023; Malmendier and Nagel, 2016). Recent evidence also indicates that labor-market news can move household expectations, including inflation expectations, more strongly than inflation news itself (Mitra and Singh, 2025). Prior dimension-reduction evidence focuses mainly on the cross section of levels: respondents who expect higher inflation also tend to expect higher unemployment and weaker aggregate outcomes, yielding a sentiment-like factor in U.S. data (Coibion et al., 2018; Kamdar and Ray, 2024), a stable two-component structure across euro-area households (Ferreira and Pica, 2024), and quantitatively important belief heterogeneity in business-cycle models (Bhandari et al., 2025). My measure is different. It is measured within respondent over time, relative to that respondent's own average answers, and it captures tightness of comovement rather than a good-times-versus-bad-times direction. In the SCE, for example, unemployment-higher and stock-price-higher probabilities move together rather than oppositely. Panel methods for grouped heterogeneity make discrete types a natural starting point (Bonhomme and Manresa, 2015). I require a type label to replicate across independent pieces of the same respondent history. Because discrete classes do not pass that test in the 12-wave SCE panel, the main result is a continuous signed-comovement ranking, avoiding labels that the public panel cannot support.

**Expectations, behavior, and cross-survey evidence.** The behavioral motivation comes from work linking household inflation expectations to spending and intertemporal choice (Armantier et al., 2015; Bachmann et al., 2015; Burke and Ozdagli, 2023; Coibion et al., 2023; Crump et al., 2022; Dräger and Nghiem, 2021; Duca-Radu et al., 2021). Experiments show that exogenous shifts in macroeconomic expectations can move spending and saving plans (Roth and Wohlfart, 2020); the exercise here is instead reduced-form predictive content, not an identified consumption response to an inflation-expectations shock. The ECB Consumer Expectations Survey provides the cross-survey comparison because it is a large repeated public panel in a different geography and macroeconomic setting (Bańkowska et al., 2021; Duca-Radu et al., 2021; Georgarakos and Kenny, 2022). Its role is to ask whether within-person macro comovement appears when question wording, response scales, countries, and time periods change.

### 3.3 Conceptual Framework: Why Joint Expectation Movements Matter

The four expectations in the SCE benchmark form a small, household-facing macro block. Inflation expectations describe the price environment and the purchasing-power side of the household problem. The probability that unemployment will be higher captures labor-market risk and real-activity conditions. The probability that saving-account rates will be higher captures a household-facing nominal return, not the policy rate itself but a rate that consumers plausibly observe. The probability that stock prices will be higher reflects asset-market and financial-condition expectations. A respondent who answers all four questions repeatedly is reporting a small panel of beliefs over prices, labor risk, nominal returns, and asset-market conditions.

Standard macro logic gives good reasons to study relationships among these expectations. Price and labor-market expectations are linked in Phillips-curve and aggregate demand-supply reasoning, but the sign need not be unique in household data. Higher inflation can be read as demand pressure and stronger activity, as cost-push pressure and weaker real conditions, or as a policy problem that will eventually tighten financial conditions and weaken labor markets. Inflation and nominal-rate expectations are linked by Taylor-rule and Fisher logic, but the SCE observes expected saving-account rates, so the relevant measure is the rate environment households actually face. Consumption Euler-equation logic makes real interest rates relevant for intertemporal spending, but the spending implication depends on income risk, borrowing constraints, and whether respondents associate inflation with nominal-income growth, price pressure, or tighter policy. Asset-pricing logic links rates and stock prices through discount rates, growth expectations, risk premia, and financial stress. These mechanisms make joint expectations meaningful, but they do not imply that one correlation matrix should be imposed on all households.

The same logic also explains the behavioral prediction. A one-variable inflation expectation can mix several forces. If a household treats higher inflation as a sign of stronger nominal demand, expected spending may rise with inflation expectations. If the same household also expects unemployment risk or saving rates to rise in those months, those linked expectations can push spending plans in the other direction. The observed inflation-spending association can therefore be flatter for households whose inflation expectations sit inside a more tightly connected expectation system. No single force need dominate; the slope researchers estimate for one expectation depends on what else tends to move with that expectation for the same household.

Three forces are especially relevant for the empirical design. Standard macro interpretations can imply different signs, since demand news, cost-push pressure, expected tightening, and financial stress need not move inflation, unemployment risk, rates, and stock-price expectations in the same way. Broad perceived macro change or uncertainty can raise several directional probabilities at once without being an optimism-pessimism factor. And common probability-scale behavior can mechanically create same-direction movement across probability answers. Because of that ambiguity, I start from all six within-person correlations and validate which summaries are stable and remain informative after levels and response style. Table C.6 in Appendix C.2 reports illustrative sign patterns.

This framework also clarifies why I start from correlations. Let $x_{it}$ denote respondent $i$'s vector of reported expectations in wave $t$. I do not observe the respondent's information set, perceived shocks, or subjective causal graph; I observe only repeated realizations of $x_{it}$. Within-respondent correlations ask whether two components of this vector are high in the same months relative to that respondent's own average answers. They are scale-free, which matters because inflation is measured in percentage points while the other SCE inputs are probabilities. They do not identify direction of causality. In particular, the same within-respondent correlation can arise from different subjective causal models: the measure records whether and how strongly two reported expectations move together, not which belief drives which or what model generates the comovement.

For the primary SCE measure, probability signed comovement, this distinction matters empirically. A positive correlation among unemployment-higher, saving-rate-higher, and stock-price-higher probabilities can reflect perceived macro change, attention to broad news, uncertainty, or probability-scale behavior. Directional mechanisms such as inflation-driven tightening or financial stress instead imply mixed signs across pairs, which is part of why the positive unemployment–stock-price link is informative. The empirical task is to show which correlation patterns are systematic and reliable enough to build on, and which should be treated as weaker signals.

Later outcomes are informative for the same reason. If a household decision or another reported expectation loads on several macro beliefs at once, then the relationship among those beliefs can matter beyond each belief's level. This motivates asking whether a validated comovement measure predicts later expectation dynamics or expected spending growth. The spending test is especially informative because it distinguishes level prediction from slope prediction. The measure may add little to expected spending levels and still change how a given inflation expectation maps into spending when offsetting labor-market or interest-rate expectations move with it.

Formally, let $z_{it}$ denote the same expectations standardized within respondent, and suppose another reported expectation or outcome $\tilde{y}_{it}$, demeaned within respondent, is associated with the vector through the descriptive approximation $\tilde{y}_{it} = \lambda_i' z_{it} + u_{it}$, where the coefficients $\lambda_i$ summarize a reduced-form mapping rather than structural parameters. A regression of $\tilde{y}_{it}$ on one component $z_{ijt}$ alone then recovers

$$\mathrm{Cov}_i(z_{ijt}, \tilde{y}_{it}) = \lambda_{ij} + \sum_{k \neq j} \lambda_{ik}\, \rho_{i,jk} + \mathrm{Cov}_i(z_{ijt}, u_{it}), \tag{3.1}$$

where $\rho_{i,jk}$ is the within-respondent correlation between expectations $j$ and $k$. Two households with the same average inflation expectation therefore need not exhibit the same one-variable inflation slope when inflation comoves differently with their unemployment, interest-rate, or asset-market expectations. The decomposition assigns no causal direction; Appendix C.3 states its assumptions and limits. The empirical spending test below implements this implication in a reduced-form cross-sectional version: it asks whether a validated respondent-level descriptor of expectation-system integration moderates the association between average inflation expectations and average expected spending growth.

## 3.4 Data and Measurement Design

### 3.4.1 SCE data and benchmark sample

The main data source is the public NY Fed Survey of Consumer Expectations. The SCE is a rotating household panel with repeated monthly observations and a broad set of macroeconomic, financial, and personal expectations.[1] I build the analysis from the public SCE microdata files. The raw SCE files used here contain 180,268 respondent-month observations for 23,886 respondents over 143 survey months. Appendix C.1 summarizes the survey items used to construct the SCE and ECB CES variables.

The benchmark sample is much smaller than the raw file by design. The raw file contains every observed respondent-month, including respondents who enter late, leave early, miss one of the benchmark questions, or have more than the desired number of complete waves. The measurement exercise instead needs a balanced respondent history: to compute and validate a within-household correlation, each retained respondent must answer the same four benchmark expectations repeatedly. I therefore begin with respondent-months that have non-missing values for the four benchmark measurement inputs, then retain respondents with exactly 12 complete survey waves. Respondents with more than 12 complete waves are excluded from the benchmark rather than truncated to their first 12; this choice is immaterial for the results, since keeping all respondents with at least 12 complete waves and using their first 12 gives nearly identical measures (Table C.15). The common 12-wave length keeps the construction uniform: each retained respondent contributes one 12-wave panel, all primitive correlations use the same time-series length, and split-half validation always compares two six-wave pieces of the same measure.

[1] The SCE elicits expectations from one respondent per household; I use household and respondent interchangeably throughout.

Table C.1 reports the sample construction. The 12-wave benchmark contains 6,659 respondents and 79,908 respondent-month observations. The raw-Pearson common-support sample then removes respondents with zero within-person variance in at least one benchmark input, because a within-household correlation is undefined when either variable is constant over the 12 waves. This second restriction is small: it removes 109 respondents and leaves 6,550 respondents for the benchmark measurement tables.

The measurement sample is defined from the four expectation inputs alone; spending does not enter the construction of the comovement indices. The 12-wave rule counts complete input observations and does not require calendar consecutiveness; consecutive panels are reported as a robustness check. A common respondent set is maintained across the main measurement tables so that index comparisons are not driven by changing samples. All statistics are unweighted.

Alternative sample windows and a panel-length audit leave the main pattern and the reliability hierarchy essentially unchanged (Table C.15, Table C.16; Appendix C.7).

Table C.17 reports a compact retained-versus-excluded comparison in the appendix. The benchmark respondents are not a random draw from all public SCE respondents: those with fewer than 12 complete benchmark waves have fewer observations by construction and somewhat higher spending-tail and focal-response shares, while respondents with more than 12 complete waves are a small group.

### 3.4.2 SCE variables and transformations

The benchmark measurement design uses four 12-month expectations: the public density-implied one-year inflation mean and the reported probabilities that the U.S. unemployment rate, average saving-account interest rates, and U.S. stock prices will be higher 12 months ahead. The inflation mean is preferred because it uses the respondent's full reported distribution and avoids making the open-ended point forecast the benchmark input.[2]

The three probability answers share a 0 to 100 percent-chance scale. Values outside that range are set to missing before constructing complete-input panels. No winsorization or trimming is applied to the four benchmark measurement inputs in the raw-Pearson construction; tail and scale concerns are handled through the density-implied inflation mean, the finite-correlation support rule, and nearby measurement variants. Appendix C.1 reports input moments and construction details.

The behavioral outcome used later is the public signed expected spending-growth field.[3] Spending tail rules apply only to the behavioral section; the comovement measure is unaffected.

### 3.4.3 ECB CES comparison sample

The ECB Consumer Expectations Survey provides the cross-survey comparison. The CES is a monthly online panel of euro-area consumers (European Central Bank, 2024). The raw monthly files used here cover 69 survey rounds, 11 countries, 1,198,212 respondent-month observations, and 137,181 respondents. I use the ECB CES after the SCE measure has been defined, to ask

[2] The signed point forecast is kept as a robustness input. Appendix C.1 maps these measures to the public SCE fields.

[3] Appendix C.1 maps this measure to the public SCE field and records the sign convention.

whether within-person macro comovement appears in a different survey environment.

The ECB design differs from the SCE benchmark in two ways that matter for interpretation. First, the primary ECB panel is a first-12 design: I retain respondents with at least 12 complete monthly observations for the ECB inputs and use their first 12 complete observations. This gives 41,200 first-12 respondents before finite-correlation restrictions. An exact-12-only ECB sample would leave only 1,749 respondents and would discard the panel depth available in the CES. The first-12 rule preserves the common 12-observation measurement length while adapting the sample rule to a survey with many longer respondent histories. Second, the ECB variables are not exact SCE analogues. The four inputs are expected inflation, expected unemployment-rate change, expected mortgage rates, and expected economic weakness.[4]

Because the ECB variables are open-ended percent responses and the survey spans countries and waves, I report several tail and scaling variants. The raw Pearson version is the closest analogue to the SCE benchmark and has 38,885 respondents with finite pairwise correlations. Country-wave standardization removes country-wave location and scale differences before computing within-person correlations. Country-wave winsorization limits the influence of tails within each country-wave cell. A raw Spearman version checks robustness to a rank-based monotone transformation. There is no direct monthly ECB analogue to the SCE stock-price probability in the primary measure; home-price expectations are kept only as a secondary asset-price candidate.

### 3.4.4 Primitive correlations and signed-comovement indices

As in Section 3.3, $x_{it}$ denotes the vector of four expectations for respondent $i$ in wave $t$. For each benchmark respondent, I compute the six pairwise within-household correlations

$$\rho_{i,jk} = \frac{\sum_{t=1}^{12}(x_{ijt} - \bar{x}_{ij})(x_{ikt} - \bar{x}_{ik})}{\sqrt{\sum_{t=1}^{12}(x_{ijt} - \bar{x}_{ij})^2}\sqrt{\sum_{t=1}^{12}(x_{ikt} - \bar{x}_{ik})^2}}, \tag{3.2}$$

where $j$ and $k$ index two different expectation inputs. Positive values mean two expectations tend to be high in the same respondent-months relative to that respondent's own means. Negative values mean they tend to move in opposite directions. Values near zero mean little linear comovement within the respondent's 12-wave panel.

The benchmark uses raw Pearson correlations. I also construct Pearson correlations after survey-month standardization and Spearman correlations of the raw inputs. The month-standardized variant checks whether the measure remains after removing common survey-month shifts. The Spearman variant checks robustness to tails and monotone transformations. The raw Pearson version remains the benchmark because the measure is signed comovement computed directly from the reported expectations.

The correlation is computed separately for each respondent and each pair of inputs; there is no pooling across respondents in the primitive correlation. The cross-sectional distributions in the measurement section are distributions of respondent-level correlation estimates.

[4] Appendix C.1 maps these inputs to the public ECB CES items and records the June 2022 mortgage-rate item change.

The four reduced summaries are simple averages of the primitive correlations:

$$I_i^{\text{inflation}} = \frac{1}{3} \sum_{k \neq \text{inflation}} \rho_{i,\text{inflation},k}, \tag{3.3}$$

$$I_i^{\text{probability}} = \frac{1}{3} \left( \rho_{i,\text{unemp,rates}} + \rho_{i,\text{unemp,stocks}} + \rho_{i,\text{rates,stocks}} \right), \tag{3.4}$$

$$I_i^{\text{overall}} = \frac{1}{6} \sum_{j<k} \rho_{i,jk}, \tag{3.5}$$

$$I_i^{\text{gap}} = I_i^{\text{probability}} - I_i^{\text{inflation}}. \tag{3.6}$$

I call these signed-comovement indices because they average the correlations with their signs retained. A respondent whose expectations move tightly together receives a high value; one whose expectations move tightly in opposite directions receives a negative value. The indices measure same-direction comovement, not dependence per se.

### 3.5 SCE Measurement Results

The measurement exercise begins with the six primitive correlations. Table 3.1 reports them before any reduction to indices or components; this table shows which parts of the correlation matrix drive the results.

Table 3.1: Within-household correlations between expectations

| Correlation pair | Mean | SD | p5 | Median | p95 |
|---|---|---|---|---|---|
| Inflation / unemployment | 0.065 | 0.357 | -0.530 | 0.071 | 0.654 |
| Inflation / saving rates | 0.041 | 0.358 | -0.541 | 0.044 | 0.633 |
| Inflation / stock prices | -0.007 | 0.357 | -0.595 | -0.008 | 0.584 |
| Unemployment / saving rates | 0.221 | 0.393 | -0.453 | 0.251 | 0.823 |
| Unemployment / stock prices | 0.150 | 0.415 | -0.554 | 0.166 | 0.809 |
| Saving rates / stock prices | 0.295 | 0.391 | -0.403 | 0.327 | 0.880 |

*Note:* The table reports raw Pearson correlations computed within each exact-12 respondent on the common-support sample of 6,550 respondents.

Three patterns stand out. First, the three probability-probability correlations are all positive and substantially larger than the inflation links. Saving-rate and stock-price probabilities have the highest mean correlation, 0.295. Unemployment and saving-rate probabilities have mean correlation 0.221, and unemployment and stock-price probabilities have mean correlation 0.150. The unemployment-stock-price correlation is especially telling. If this were a simple good-times versus bad-times sentiment index, a higher probability of unemployment and a higher probability of stock prices would naturally move in opposite directions. They do not on average. Probability expectations themselves tend to move together within many households.

The three probability correlations capture a respondent-level tendency to move several probability assessments together in the same survey months, relative to the respondent's own averages. Section 3.6 returns to what could generate this tendency and tests the response-style channel directly. Whatever the mechanism, public SCE panel answers contain a systematic household-level probability-comovement dimension.

Second, inflation is much less integrated with the other expectations in the benchmark panel. The mean inflation-unemployment correlation is 0.065, the mean inflation-saving-rate correlation is 0.041, and the mean inflation-stock-price correlation is slightly negative at -0.007. Inflation expectations are central in the household-expectations literature, but in this 12-wave SCE design the density-implied inflation mean does not move strongly with the three probability expectations within respondent. Inflation may still matter for levels, for forecast errors, and for behavior, but its within-household comovement with these other macro expectations is limited in this design. From the perspective of policy communication, this weak link is itself a finding: of the four expectations, inflation is the one whose movements contain the least information about the rest of the household's macroeconomic beliefs in these panels.

Given the contrast between the probability block and the inflation links, the main measure is narrower than a general macro optimism factor: it captures how respondents move their probability expectations together. That concentration disciplines interpretation, since a purely mechanical tendency to give uniformly high or low answers in a month would not predict a signal specific to the probability-format inputs.

Appendix C.6 checks two mechanical explanations for the weak inflation links. The conclusion is unchanged when the inflation input is put on a 0–100 percent-chance scale or when duplicate-elicitation evidence is used to assess measurement-noise attenuation.

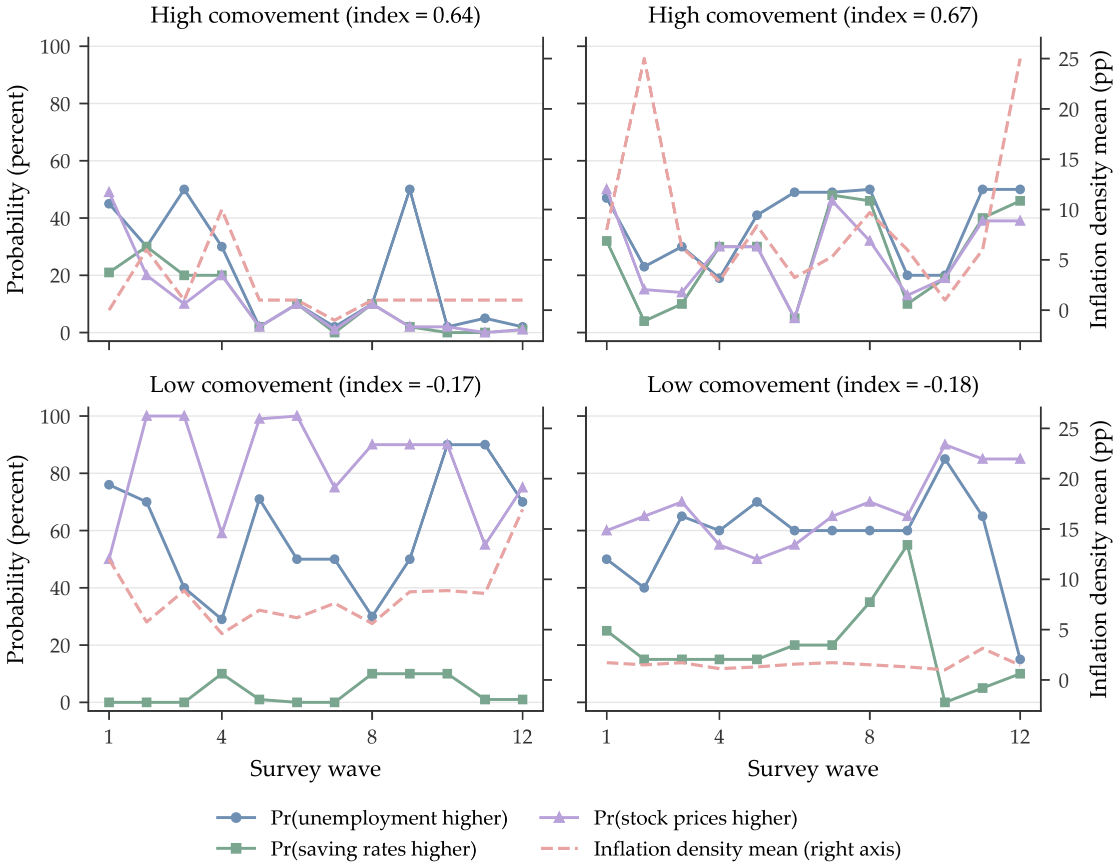
Third, there is substantial cross-household heterogeneity. For every primitive correlation, the fifth and ninety-fifth percentiles span negative and positive values. Some households answer as if macro probabilities rise and fall together; others answer as if selected expectations move in opposite directions; others show little stable comovement. The averages are only the center of the distribution of interest: I focus on the full cross-sectional distribution of household-specific correlation patterns. The tails of that distribution are economically meaningful even though individual correlations are noisy; the measurement question is whether this ordering of households is stable enough, and distinct enough from levels and response styles, to be informative.

Figure 3.1 makes this heterogeneity concrete with raw answer paths for four respondents from the high and low tails of the probability signed-comovement distribution. The figure gives a visual anchor for the index: it shows the kind of within-household movement summarized by the ranking before the validation checks quantify its stability.

Table 3.2 summarizes the reduced signed-comovement indices built from those primitive correlations. In the raw-Pearson benchmark the probability index averages 0.222, roughly seven times the inflation-link mean of 0.033; the overall and gap indices fall in between.[5] For a respondent at the mean correlation of 0.222, two probability answers share only about 5 percent of their within-respondent variance, so the typical respondent is far from moving all three probabilities in perfect alignment. The relevant variation is the spread around that mean, from -0.233 at the fifth percentile to 0.766 at the ninety-fifth: respondents near the top of the ranking move the three probabilities nearly in parallel, while respondents near the bottom move

[5]This benchmark mean is computed on the 6,550-respondent common-support sample. The robustness checks in the appendix recompute the same index on row-specific common samples, which differ from this benchmark by the additional data each check requires (a second inflation elicitation, valid probabilistic bins, or an available density IQR). The resulting differences in the third decimal reflect sample composition, not changes to the index.

Figure 3.1: Example respondents: four expectations over 12 survey waves




*Note.* Each panel shows one benchmark respondent's reported answers across that respondent's 12 complete waves, which need not be calendar-consecutive. Solid lines (left axis) are the three probability inputs on the 0–100 scale; the dashed line (right axis) is the density-implied inflation mean in percentage points, on a common scale across panels. Respondents are selected by a deterministic rule: among common-support respondents with a probability signed-comovement index percentile rank between 0.88 and 0.92 (top row) or between 0.08 and 0.12 (bottom row), the figure shows the two respondents whose average within-person standard deviation across the three probability inputs is closest to the sample median. Panel headers report each respondent's index.

them independently or in opposite directions. The reduced table captures the same fact as the primitive table: the reliable structure is concentrated in the probability block.

Table 3.2: SCE signed-comovement measures

| Measure | *N* | Mean | Median | p5 | p95 | Month-std. mean | Spearman mean |
|---|---|---|---|---|---|---|---|
| Probability | 6,550 | 0.222 | 0.198 | -0.233 | 0.766 | 0.224 | 0.211 |
| Inflation | 6,550 | 0.033 | 0.032 | -0.364 | 0.435 | 0.028 | 0.031 |
| Overall | 6,550 | 0.128 | 0.100 | -0.156 | 0.496 | 0.126 | 0.121 |
| Probability − inflation | 6,550 | 0.189 | 0.144 | -0.340 | 0.882 | 0.196 | 0.180 |

*Note:* Rows summarize the exact-12 SCE common-support sample. The probability index is the primary measure; it averages the three within-household correlations among unemployment, saving-rate, and stock-price expectations. The month-standardized and Spearman columns report nearby measurement choices.

The probability index is the most interpretable summary of the underlying correlation pattern. It uses exactly the three links where the mean comovement is largest, treats them symmetrically, and reports the weaker inflation links separately rather than averaging them away. The gap index serves as a contrast, but the level of probability comovement has the most direct interpretation.

The ranking also appears under alternative measurement choices. The month-standardized and Spearman columns of Table 3.2 show the same gap between the probability and inflation indices, and the sample-window and differenced checks in the appendix preserve the probability-versus-inflation ordering.

Figure 3.2 shows the main distributional contrast. The probability index is shifted to the right of the inflation-link index, and both distributions remain wide. The figure makes clear that households vary widely in how their macro expectations move together over time. Figure C.2 in the appendix reports the full set of reduced signed-comovement summaries.

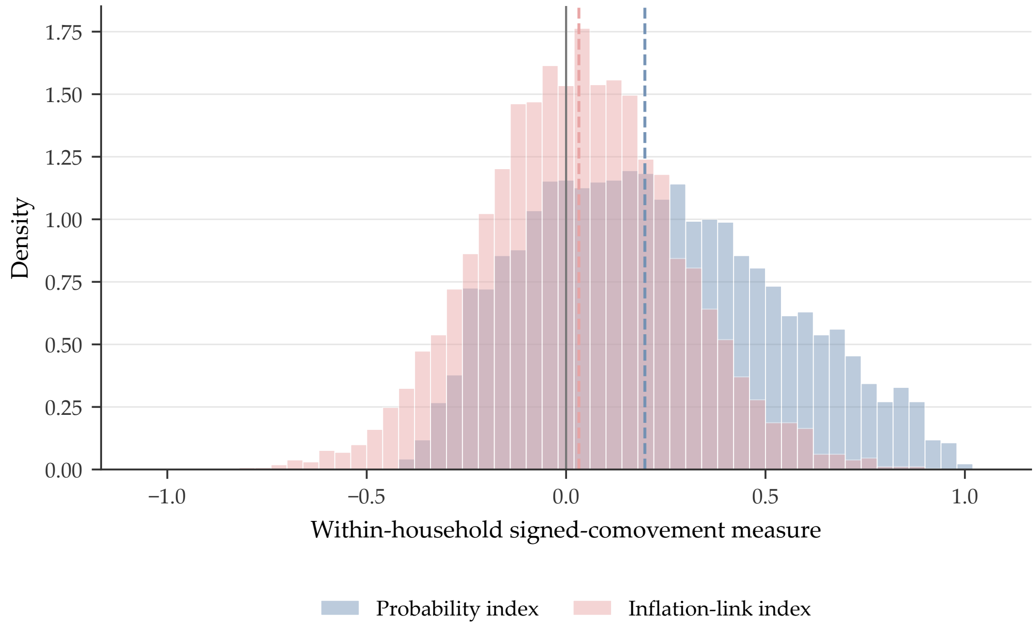
The evidence points to a continuous ranking. The probability signed-comovement index is the baseline measure because it tracks the reliable part of the six-correlation structure. PCA confirms this choice. In Table C.9, the first component loads primarily on the three probability-probability links and explains 30 percent of the variance in the six standardized correlations; its correlation with the index is 0.959 in the raw-Pearson benchmark. The index gives almost the same empirical ordering as the first component while remaining easier to explain. Its components can be inspected variable by variable; a component score is more compact but less directly tied to the survey questions.

Clustering is an alternative summary, but class assignments do not replicate across independent halves of the same respondent histories (Section 3.6, Appendix C.10).

## 3.6 SCE Validation

The validation addresses three basic questions: whether the observed links are stronger than what would appear if each respondent's answers were randomly reshuffled over time, whether cross-household rankings remain detectable when each respondent's history is split into shorter pieces, and whether the ranking is more than a relabeling of average expectations, demographics, numeracy, timing, or response style. Table 3.3 and Table 3.4 report these checks.

Figure 3.2: Probability and inflation-link comovement in the SCE




*Note.* The figure plots density-normalized histograms for the 12-wave raw-Pearson probability and inflation-link signed-comovement indices in the common-support sample. Dashed vertical lines mark medians. Histogram ranges include all observed values for the plotted measures.

Table 3.3 compares the reduced indices to a shuffled-time benchmark. For each respondent, I randomly reshuffle the survey months of each expectation input, recompute the same primitive correlations, and then recompute the same reduced indices. This preserves the respondent's distribution of answers for each variable, including tails, focal answers, and low-variation series, while breaking the month-by-month alignment across variables. The benchmark tests whether the observed correlations exceed what each respondent's own answer distributions produce mechanically.

The observed probability index mean of 0.222 is far outside the null range, whose 95th percentile is 0.003. The overall index and the probability-minus-inflation gap are also far from the null. Even the smaller inflation-link mean of 0.033 is above the null range. The basic fact for the rest of the validation is that repeated SCE answers contain within-person temporal alignment across expectation variables, not only marginal answer distributions.

Table 3.3: Reduced SCE measures against within-person time-permutation nulls

| Measure | Observed | Null p5 | Null p50 | Null p95 |
|---|---|---|---|---|
| Probability | 0.222 | -0.003 | 0.000 | 0.003 |
| Inflation | 0.033 | -0.003 | 0.000 | 0.003 |
| Overall | 0.128 | -0.003 | 0.000 | 0.002 |
| Probability − inflation | 0.189 | -0.006 | 0.000 | 0.004 |

*Note:* Null entries are computed at the iteration level from the same feature averages that define the reduced signed-comovement measures, using 100 permutation iterations. The permutation breaks within-person time alignment across expectation variables while preserving each respondent's marginal series.

Table 3.4 then asks whether the household ranking is stable. The main check repeatedly divides each respondent's 12 waves into two six-wave halves, recomputes the measure in each half, and asks whether households keep similar ranks across the two pieces. This is a demanding test because each half contains only six observations, so each half-sample correlation is noisy. The table also reports first-last and odd-even reliability, a descriptive control $R^2$, and a Spearman-Brown half-to-full adjustment, the standard projection from reliability in two six-wave half-measures to reliability for the full 12-wave measure. The appendix gives the exact formulas and additional reliability checks; the main text uses the table to decide which measure is stable enough to build on.

The control blocks are descriptive reducibility checks covering average expectation levels, focal response behavior, timing and panel-gap variables, demographics, education, income, and numeracy. Appendix C.4 lists the full variable set.

Table 3.4: Validation hierarchy for SCE comovement measures

| Measure | Random split | Spearman–Brown | First/last | Odd/even | Controls $R^2$ |
|---|---|---|---|---|---|
| Probability | 0.394 | 0.565 | 0.283 | 0.436 | 0.286 |
| Projected PC1 | 0.369 | 0.539 | 0.259 | 0.412 | 0.268 |
| Overall | 0.277 | 0.433 | 0.174 | 0.317 | 0.194 |
| Probability − inflation | 0.255 | 0.406 | 0.148 | 0.294 | 0.160 |
| Inflation | 0.114 | 0.204 | 0.011 | 0.152 | 0.012 |
| Projected PC2 | 0.133 | 0.234 | 0.026 | 0.173 | 0.027 |

*Note:* Random-split columns average 200 respondent-specific random 6/6 splits; the Spearman–Brown column applies the standard half-to-full-length adjustment. Controls R-squared is the descriptive projection on expectation levels, focal response indicators, timing variables, demographics, education, income, and numeracy where available.

Probability signed comovement is the primary measure. Its mean random-split reliability is 0.394, which projects to 0.565 for the full 12-wave measure under the Spearman-Brown adjustment. This means the measure is noisy at the individual level but informative for comparing households in the cross section. The odd-even reliability points in the same direction, while the lower first-last reliability is consistent with a measure that is partly stable and partly exposed to true change over the respondent's panel window. Projected PC1 is similar and offers no reliability gain over the simple probability average. Overall comovement and the probability-minus-inflation gap are secondary. Inflation signed comovement is present on average, but its split-half reliability is too low to support the same household-ranking interpretation.

Observed characteristics and response behavior explain a meaningful but incomplete share of the ranking. A full block of expectation levels, focal response measures, timing variables, demographics, education, income, and numeracy explains 28.6 percent of probability signed comovement; relative to the index's reliability, the controls capture roughly half of the systematic component. The appendix reports the block decomposition, signed demographic gradients, probability-placebo checks, and timing diagnostics. Levels and respondent characteristics matter, but the probability index remains largely distinct from non-macro probability answering and appears in pre-pandemic, pandemic/reopening, and 2022-onward respondent windows.

The validation evidence also informs mechanism. The SCE inflation-density question asks respondents to assign probabilities to future inflation outcomes, so the interquartile range of that distribution measures the width of reported inflation beliefs. Respondents with wider average inflation distributions have higher probability comovement: the cross-sectional correlation between average density IQR and the probability index is 0.234. Within respondent, months with wider inflation distributions are also months in which the three probability inputs tend to be higher, with a mean correlation of 0.053 against a shuffled-time null whose 95th percentile is 0.004. This link rises across the comovement ranking, from 0.004 in the bottom quartile to 0.134 in the top quartile. The pattern fits the interpretation that some high-comovement households sometimes view the macro environment as broadly unsettled, so several outcomes feel more likely to move at once. This interpretation is related to evidence that survey respondents can be cognitively uncertain about their own judgments (Enke and Graeber, 2023), although the SCE measure here captures reported inflation uncertainty rather than cognitive uncertainty directly. Because the IQR measures inflation uncertainty rather than overall macro uncertainty, Appendix C.8 reports residualization and full diagnostics rather than treating the IQR as a decomposition of the index.

The same validation standard governs discrete types. The appendix clustering diagnostics in Table C.30 show that full-sample k-means partitions are stable across random starts and bootstrap resamples, so the six-correlation space has stable full-sample structure. The split-half adjusted Rand indices are near zero, however, so individual class labels do not replicate when the same respondent is measured with two independent six-wave halves.

The validation evidence supports a continuous measure of probability-expectation comovement. Observed levels and characteristics explain part of the measure while leaving a large residual, and the placebo checks distinguish it from generic probability answering. Inflation-link comovement is present on average but lacks the split-half reliability needed for the same household-ranking interpretation. Discrete types are exploratory summaries, not stable respondent-level categories in this 12-wave design. I use probability comovement as the primary SCE measure and treat the other summaries as supporting evidence. Table C.7 in Appendix C.5 summarizes this hierarchy and the corresponding limits on interpretation.

**Interpreting the validated measure.** The probability signed-comovement index measures household-level variation in how three probability expectations move together within respondent. A high value means that, relative to that respondent's own average answers, unemployment, saving-rate, and stock-price probabilities tend to rise and fall together across the 12 survey waves. A low value means that these probability expectations tend to move separately or in opposite directions. Because unemployment-higher and stock-price-higher probabilities move together rather than oppositely, a high index value signals strong same-direction links among probability expectations rather than a directional view of the economy. Respondents whose probability answers move tightly in opposite directions appear in the low tail of the index, as shown by the primitive-correlation distributions. The sign-agnostic starting point of Section 3.3 is kept at the level of the six primitive correlations; the positive average sign of the probability block is an empirical result, and it is what licenses a signed average as the summary.

I interpret the index descriptively: it captures interdependence among reported expectations that remains after common-month shocks, response style, and expectation levels are taken into account. The detailed interpretation checks in Appendix C.8 include non-macro probability placebos, timing diagnostics, links to inflation-density uncertainty, and same-direction revision evidence. They support reading the index as a stable household descriptor, without assigning it to a single causal or psychological mechanism.

### 3.7 Evidence from the ECB Consumer Expectations Survey

The ECB Consumer Expectations Survey provides a stringent comparison because it asks different questions, uses open-ended percent responses instead of the SCE probability scale, covers a different macroeconomic environment, and has no monthly stock-price analogue in the primary measure. The exercise tests portability of the broader idea—within-person macro expectations comove in repeated household panels—rather than attempting to replicate the exact SCE probability block.

The ECB measure is an adverse macro-comovement index. It averages within-person correlations among expected inflation, expected unemployment-rate increases, expected mortgage rates, and expected general economic weakness. The signs are chosen so that higher values point toward a less favorable macroeconomic assessment: higher expected inflation, higher expected unemployment, higher expected mortgage rates, and weaker expected activity. Higher values of the index mean that adverse macro expectations move together within the respondent's first 12 usable CES observations.

This construction changes the interpretation relative to the SCE probability index. In the SCE, the dominant signal is a probability-format block that includes stock-price probabilities and is distinct from an adverse-sentiment index; in the CES, the available monthly variables make an adverse-comovement measure more appropriate. The comparison asks whether repeated household expectations in a different public panel contain systematic within-person macro comovement once the variables are put on a coherent sign convention.

The adverse macro-comovement index has a raw-Pearson mean of 0.084 and median of 0.060 among 38,885 finite respondents (Table 3.5). The fifth percentile is negative and the ninety-fifth percentile is 0.418, so the result has the same qualitative shape as the SCE result: a cross section centered above zero with wide dispersion. The mean is well below the SCE probability-index mean of 0.222, as one would expect from open-ended percent responses, imperfect variable analogues, and noisier inputs.

The table shows a stable pattern across scale and tail choices. The raw, country-wave-standardized, winsorized, and Spearman variants all have positive means well above shuffled-time benchmarks. Behind the index, the largest single primitive link is between unemployment and economy-weakness expectations, with mean 0.191, while the remaining pairs are smaller but also centered above zero (Table C.26). Country and timing indicators explain little of the variation, and replacing open-ended inflation with a midpoint-implied mean from probabilistic inflation bins barely changes the adverse-comovement mean. Appendix C.9 reports these diagnostics and the distribution plots.

Table 3.5: ECB CES evidence on adverse macro comovement

| Variant | $N$ | Mean | Median | p5 | p95 | Random | First/last | Odd/even | Null p95 | Timing $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Raw Pearson | 38,885 | 0.084 | 0.060 | -0.168 | 0.418 | 0.162 | 0.048 | 0.219 | 0.001 | 0.011 |
| Country-wave z | 41,200 | 0.072 | 0.053 | -0.165 | 0.381 | 0.127 | 0.037 | 0.156 | 0.001 | 0.008 |
| Country-wave winsor | 38,896 | 0.085 | 0.061 | -0.167 | 0.420 | 0.164 | 0.048 | 0.220 | 0.001 | 0.011 |
| Spearman | 38,885 | 0.083 | 0.063 | -0.160 | 0.400 | 0.164 | 0.048 | 0.226 | 0.001 | 0.015 |
| Country-wave z, common support | 38,885 | 0.073 | 0.053 | -0.165 | 0.379 | 0.120 | 0.037 | 0.146 | 0.001 | 0.007 |

*Note:* The ECB CES measure uses harmonized analogues for inflation, unemployment, mortgage rates, and general economic weakness. It is cross-survey evidence, not a one-for-one replication of the SCE probability index. Timing R-squared is from country and window-midpoint cohort indicators. The ECB estimates use 200 random-split iterations and 100 null iterations. Country-wave z-scoring can turn a constant raw answer series into a varying relative-position series, which is why the standardized variant is finite for more respondents than the raw variant; the common-support row recomputes the standardized variant on the raw-Pearson finite sample so the transformation is evaluated on a fixed respondent set.

The CES evidence is strongest as distributional portability. Random-split reliability is positive in every variant, but it is lower than in the SCE: 0.162 in the raw-Pearson row and 0.120 to 0.164 across the other rows. First-half versus second-half reliability for the adverse index stays below 0.05, although odd-even reliability is higher. The CES reveals the same distributional pattern at a smaller mean, with weaker respondent-level ranking than the SCE probability index.

Adverse macro-comovement also appears outside the SCE: it survives country-wave and tail treatments, lies well above a within-person shuffled-time benchmark, and shows that linked expectation structures are not confined to one U.S. questionnaire.

## 3.8 What the Measure Predicts

The final empirical section asks what the validated measure predicts. This evidence is secondary to the measurement exercise, but it is important for economic interpretation: if expectation systems matter, they should change how familiar expectation-behavior relationships are read. I use the measure in three ways. Early comovement predicts later organization inside the expectation system. Direct prediction of expected spending growth and personal-risk expectations is limited. And I test whether the same inflation expectation can map differently into expected spending depending on the surrounding expectation system, the reduced-form behavioral implication motivated by the framework. This slope test, rather than direct spending-level prediction, is the main behavioral result. Appendix C.11 reports the spending sample construction, secondary regressors, robustness checks, and personal-risk margins.

### 3.8.1 Predicting later expectation dynamics

The first predictive test remains within the expectation system. I recompute probability signed comovement from waves 1–6 only and ask whether it predicts expectation-system organization in waves 7–12, conditional on first-half expectation levels, focal-response and timing controls, demographics, education, income, and numeracy. The timing separation matters: the waves used to construct the regressor never enter the outcomes. This specification is more demanding than the first-last reliability statistic in Table 3.4: it conditions on the full first-half control block and adds the same-direction revision outcome, which the reliability check does not cover.

Comovement measured in the first half of the panel persists into the second half. A one-standard-deviation increase in first-half probability comovement predicts 0.066 index points more later probability comovement, against an outcome mean of 0.200 (standard error 0.005), and a 3.6 percentage-point higher share of later same-direction probability revisions, against an outcome mean of 0.599 (standard error 0.003; Table 3.6). Households with higher early comovement continue to move their probability expectations together, and to revise them in the same direction, months later.

Table 3.6: First-half comovement and later expectation dynamics

| Variable | (1) | (2) |
|---|---|---|
| Signed comovement | 0.066*** | 0.036*** |
| | (0.005) | (0.003) |
| Controls | Yes | Yes |
| Observations | 5,931 | 6,361 |
| $R^2$ | 0.138 | 0.083 |
| Mean dep. var. | 0.200 | 0.599 |

*Note:* The regressor is standardized probability signed comovement measured from waves 1–6. The dependent variable is later probability comovement in column (1) and later same-direction probability revisions in column (2), both measured from waves 7–12. All columns include first-half expectation levels, focal-response and timing controls, demographics, education, income, and numeracy. Heteroskedasticity-robust (HC1) standard errors are in parentheses. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

### 3.8.2 Direct spending and personal-risk prediction

The direct spending-level test is a benchmark, not the main behavioral implication. The slope decomposition in Equation (3.1) motivates heterogeneity in inflation-spending slopes rather than a direct level effect of comovement on spending; Section 3.8.3 turns to the slope implication.

The outcome is the public signed expected spending-growth field. I estimate respondent-level average-spending regressions, respondent-month panel regressions with survey-month fixed effects, and a first-half-to-later-spending design. All rows use the full control block: expectation levels, response-style and timing measures, demographics, education, income, and numeracy.

The direct association is essentially zero. In the full-control respondent-level specification, a one-standard-deviation increase in probability comovement is associated with -0.017 percentage points of average expected spending growth, with standard error 0.102 and $p = 0.869$ (Table C.32). The winsorized, panel, and first-half-to-later-spending rows are similarly small and imprecise; Appendix C.11 reports the table.

The null is economically small, not only statistically insignificant. The reliability adjustment reported in Appendix C.11 would not make this direct spending association meaningful. With expectation levels and response styles controlled, the probability-comovement index adds little to expected spending levels. The personal-risk margins are similar: the index does not meaningfully predict expected income growth, expected credit access, or debt-payment risk; the negative job-loss association among eligible employed respondents remains unresolved and is reported in Appendix C.11.

### 3.8.3 Belief-system structure and the inflation–spending slope

The relevant behavioral test is about slopes. A one-variable inflation-spending regression treats two households with the same average inflation expectation as comparable, even if one household's inflation expectation moves with unemployment risk and expected rates while the other's does not. The framework motivates the possibility that this surrounding expectation structure can change the inflation-spending slope.

The algebra in Equation (3.1) is clearest when the measured object is inflation-link comovement. In the SCE, however, inflation links are the least reliable respondent-level measure. I use the validated probability-comovement index as a lower-noise marker of how integrated the household's macro expectation system is. The interpretation is deliberately system-level: the interaction asks whether households whose macro probabilities move tightly together attach a different spending meaning to a given inflation expectation. The probability block is a marker of organization, not the channel being separately identified.

The test interacts standardized probability signed comovement with standardized average inflation expectations in the respondent-level spending specifications above, with both main effects and the full control block (Table 3.7). The interaction is negative and significant in both outcomes: −0.316 (standard error 0.115, $p = 0.006$) for trimmed spending and −0.381 (standard error 0.118, $p = 0.001$) for winsorized spending. The main effects imply that, at average comovement, a one-standard-deviation higher average inflation expectation is associated with about 1.5 percentage points higher expected spending growth. Moving from the 25th to the 75th percentile of comovement lowers the implied inflation slope from 1.783 to 1.310 percentage points of expected spending growth per standard deviation of average inflation expectations in the trimmed outcome (1.885 to 1.314 winsorized), a decline of roughly one quarter. The result is a flattening, not a reversal: high-comovement households still have a positive inflation-spending association, but the slope is about one quarter smaller (Figure 3.3).

This is the main behavioral content of the measure. A standard inflation-spending slope is not only a property of the inflation expectation itself; it also depends on what other macro expectations tend to move with inflation for the same household. The negative interaction is consistent with offsetting linked expectations, such as higher unemployment risk or tighter expected rates, flattening the one-variable inflation slope. The regression does not estimate household-specific spending mappings or the individual terms in Equation (3.1). It tests the reduced-form implication that the inflation-spending association differs across households whose macro expectations are more or less tightly organized, without identifying the channels separately.

Table 3.7: Spending-slope heterogeneity by probability comovement

| Variable | (1) | (2) |
|---|---|---|
| Probability comovement | -0.030 | -0.025 |
| | (0.103) | (0.106) |
| Average inflation expectations | 1.535*** | 1.585*** |
| | (0.136) | (0.136) |
| Comovement × inflation | -0.316*** | -0.381*** |
| | (0.115) | (0.118) |
| Implied inflation slope at p25 comovement | 1.783*** | 1.885*** |
| | (0.179) | (0.183) |
| Implied inflation slope at p75 comovement | 1.310*** | 1.314*** |
| | (0.144) | (0.141) |
| Outcome | Trimmed spending | Winsorized spending |
| Controls | Yes | Yes |
| Observations | 6,549 | 6,550 |
| $R^2$ | 0.073 | 0.080 |
| Mean dep. var. | 4.360 | 4.540 |

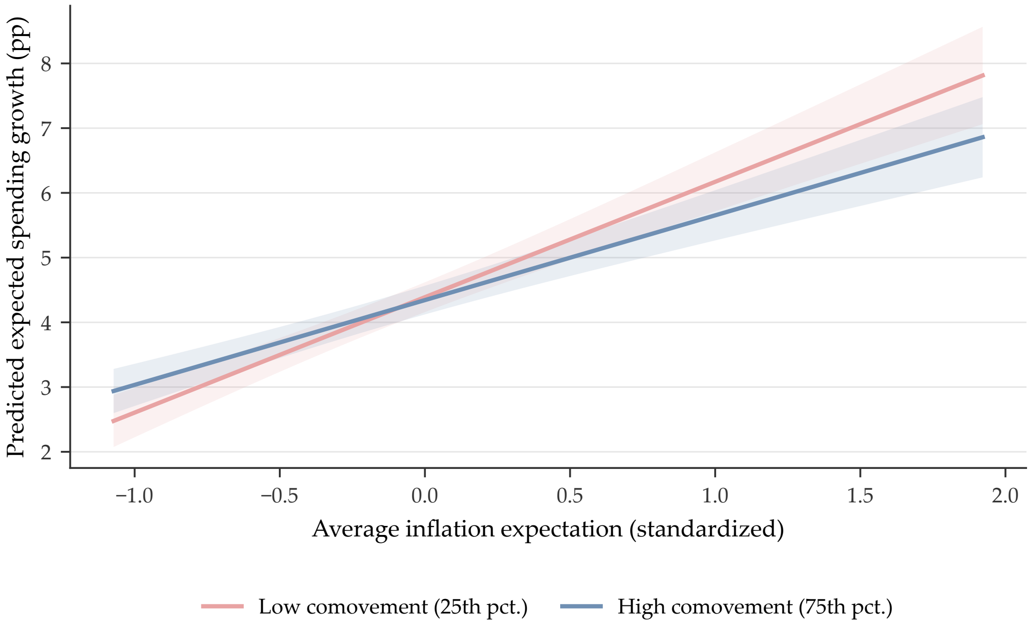
*Note:* Columns report respondent-level expected-spending-growth regressions. Coefficients are percentage points of expected spending growth per one cross-sectional standard deviation of each regressor. The implied inflation slope evaluates the inflation-expectation coefficient at the 25th and 75th percentiles of standardized comovement in the estimation sample, with delta-method standard errors from the HC1 covariance matrix. Specifications include the full control block, excluding the raw inflation-level control because its standardized version is included directly. Heteroskedasticity-robust (HC1) standard errors are in parentheses. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.37 in Appendix C.11 reports the robustness checks. The interaction remains negative when average inflation uncertainty is controlled for, when an uncertainty-inflation interaction is added, when a numeracy-inflation interaction is added, and in respondent-month specifications with survey-month fixed effects. A household bootstrap that re-draws respondents and re-runs the full estimation gives similar uncertainty around the interaction (Table C.38). Measured inflation uncertainty, numeracy, month effects, and generated-regressor uncertainty do not absorb the pattern.

The main limitation is timing. The interaction is concentrated in respondent windows centered on 2022 onward and is small and insignificant in earlier windows, while the comovement measure itself appears in all timing windows (Table C.19). This concentration is informative but limits external validity: the slope heterogeneity is most visible when inflation is salient and the baseline inflation-spending slope is steep. Whether the same behavioral gradient appears in calmer inflation regimes requires future evidence with richer outcomes, information treatments, or linked administrative data. A second limitation is measurement: replacing the density-implied inflation mean with the noisier winsorized point forecast attenuates the interaction toward zero.

The behavioral evidence is narrow but consequential: the measure has little direct spending-

Figure 3.3: The inflation–spending slope by probability comovement




*Note.* Predicted expected spending growth against the standardized average density-implied inflation expectation, from the trimmed-spending interaction in Table 3.7, evaluated at the 25th and 75th percentiles of probability signed comovement with all other controls held at their means. Because the inflation axis is standardized, the slope of each line equals the implied per-standard-deviation inflation slope reported in the text (1.78 at the 25th percentile, 1.31 at the 75th). Shaded bands are delta-method 95% confidence intervals from HC1 standard errors. The lines illustrate the estimated reduced-form interaction; they are not an identified causal slope.

level content, yet it changes how a canonical inflation-expectations coefficient should be interpreted.

## 3.9 Conclusion

Public household expectations panels contain information beyond individual expectation levels: they reveal how households connect their macro expectations over time. This paper shows that standard survey data recover that structure, that it survives extensive validation, and that it carries content for how expectations evolve and how inflation maps into behavior. The probability signed-comovement index—a within-respondent average correlation among unemployment, saving-rate, and stock-price probability expectations—is the clearest and most reliable summary of this structure in 12-wave SCE panels, and its sign pattern is not optimism or pessimism: at the top of the ranking, unemployment-risk and stock-price probabilities rise together.

The validation exercises establish a hierarchy: probability comovement is systematic, well above what noise benchmarks produce, stable in split-half validation, and only partly explained by expectation levels, demographics, and response styles. The alternatives—inflation-link comovement, principal components, and discrete household types—each fall short of that standard; the continuous probability ranking is the main measure.

A parallel adverse macro-comovement pattern appears in the ECB CES under different survey questions, geography, and macroeconomic conditions; within the SCE, early comovement predicts later expectation-system dynamics; and the inflation–spending slope varies with the structure around the inflation expectation, as the framework predicts: it falls by roughly one quarter between the 25th and 75th percentiles of comovement while remaining clearly positive. The direct spending-level signal is essentially null; the behavioral content appears in expectation dynamics and in slope heterogeneity.

The probability-comovement measure can be used in future work. It is easy to construct from repeated public survey answers and stable enough to serve as a noisy cross-sectional descriptor in the public SCE panel. Information treatments, communication experiments, and administrative-data linkages can use it as a pre-specified household-level descriptor.

More broadly, expectation surveys contain more than separate marginal expectations. For central banks, slope heterogeneity means that a single average inflation–spending association can hide systematic cross-household variation. If future designs with exogenous variation find the same pattern, inflation communication would be expected to change spending plans less among households whose macro expectations are most tightly linked. The evidence here is reduced form, and the slope heterogeneity is concentrated in the recent high-inflation period, where the mechanism should be most visible; whether the same behavioral gradient appears in calmer regimes remains open. A transparent correlation design, disciplined by noise benchmarks and reliability checks, still recovers a stable measure in existing household panels without imposing a causal graph or a typology.

# Bibliography


**Acemoglu, Daron**, "Harms of AI," Technical Report, National Bureau of Economic Research 2021.

**—, Asuman Ozdaglar, and James Siderius**, "A Model of Online Misinformation," *The Review of Economic Studies*, November 2024, *91* (6), 3117–3150.

**—, —, and —**, "AI and Social Media: A Political Economy Perspective," in Ajay Agrawal, Joshua Gans, Avi Goldfarb, and Catherine E. Tucker, eds., *The Political Economy of Artificial Intelligence*, University of Chicago Press, 2026, chapter 12, pp. 167–206.

**Alabrese, Eleonora, Francesco Capozza, and Prashant Garg**, "Politicized Scientists: Credibility Cost of Political Expression on Twitter," CESifo Working Paper 11254, CESifo 2026.

**Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow**, "The Welfare Effects of Social Media," *American Economic Review*, March 2020, *110* (3), 629–676.

**—, Matthew Gentzkow, and Lena Song**, "Digital addiction," *American Economic Review*, 2022, *112* (7), 2424–2463.

**—, —, Winter Mason, Arjun Wilkins, Pablo Barberá, Taylor Brown, Juan Carlos Cisneros, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Sandra González-Bailón, Andrew M. Guess, Young Mie Kim, David Lazer, Neil Malhotra, Devra Moehler, Sameer Nair-Desai, Houda Nait El Barj, Brendan Nyhan, Ana Carolina Paixao De Queiroz, Jennifer Pan, Jaime Settle, Emily Thorson, Rebekah Tromble, Carlos Velasco Rivera, Benjamin Wittenbrink, Magdalena Wojcieszak, Saam Zahedian, Annie Franco, Chad Kiewiet De Jonge, Natalie Jomini Stroud, and Joshua A. Tucker**, "The effects of Facebook and Instagram on the 2020 election: A deactivation experiment," *Proceedings of the National Academy of Sciences*, May 2024, *121* (21), e2321584121.

**Allen, Jennifer, Baird Howland, Markus Mobius, David Rothschild, and Duncan J. Watts**, "Evaluating the fake news problem at the scale of the information ecosystem," *Science Advances*, 2020, *6* (14), eaay3539.

**Anderson, Simon P and John McLaren**, "Media Mergers and Media Bias with Rational Consumers," *Journal of the European Economic Association*, 2012, *10* (4), 831–859.

**Andrade, Philippe, Erwan Gautier, and Eric Mengus**, "What matters in households' inflation expectations?," *Journal of Monetary Economics*, 2023, *138*, 50–68.

**Andre, Peter, Carlo Pizzinelli, Christopher Roth, and Johannes Wohlfart**, "Subjective models of the macroeconomy: Evidence from experts and representative samples," *The Review of Economic Studies*, 2022, *89* (6), 2958–2991.

__ , **Ingar Haaland, Christopher Roth, Mirko Wiederholt, and Johannes Wohlfart**, "Narratives about the macroeconomy," *The Review of Economic Studies*, 2026.

**Aridor, Guy, Rafael Jiménez-Durán, Ro'ee Levy, and Lena Song**, "The Economics of Social Media," *Journal of Economic Literature*, December 2024, *62* (4), 1422–1474.

__ , **Tevel Dekel, Rafael Jiménez-Durán, Ro'ee Levy, and Lena Song**, "Digital News Consumption: Evidence from Smartphone Content in the 2024 US Elections," 2025. Working paper.

**Armantier, Olivier, Giorgio Topa, Wilbert van der Klaauw, and Basit Zafar**, "An overview of the Survey of Consumer Expectations," *Economic Policy Review*, 2017, *23* (2), 51–72.

__ , **Scott Nelson, Giorgio Topa, Wilbert van der Klaauw, and Basit Zafar**, "The price is right: Updating inflation expectations in a randomized price information experiment," *The Review of Economics and Statistics*, 2016, *98* (3), 503–523.

__ , **Wändi Bruine de Bruin, Giorgio Topa, Wilbert van der Klaauw, and Basit Zafar**, "Inflation expectations and behavior: Do survey respondents act on their beliefs?," *International Economic Review*, 2015, *56* (2), 505–536.

**Ash, Elliott and Stephen Hansen**, "Text algorithms in economics," *Annual Review of Economics*, 2023, *15* (1), 659–688.

**Bachmann, Rüdiger, Tim O. Berg, and Eric R. Sims**, "Inflation expectations and readiness to spend: Cross-sectional evidence," *American Economic Journal: Economic Policy*, 2015, *7* (1), 1–35.

**Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky**, "Exposure to opposing views on social media can increase political polarization," *Proceedings of the National Academy of Sciences*, September 2018, *115* (37), 9216–9221. Publisher: Proceedings of the National Academy of Sciences.

**Bakshy, Eytan, Solomon Messing, and Lada A Adamic**, "Exposure to ideologically diverse news and opinion on Facebook," *Science*, 2015, *348* (6239), 1130–1132.

**Bańkowska, Katarzyna, Ana Maria Borlescu, Evangelos Charalambakis, Antonio Dias da Silva, Davide Di Laurea, Maarten Dossche, Dimitris Georgarakos, Juha Honkkila, Neale Kennedy, Geoff Kenny, Aleksandra Kolndrekaj, Justus Meyer, Desislava Rusinova, Federica Teppa, and Veli-Matti Törmälehto**, "ECB Consumer Expectations Survey: An overview and first evaluation," Technical Report 287, European Central Bank 2021.

**Baron, David P**, "Persistent Media Bias," *Journal of Public Economics*, 2006, *90* (1-2), 1–36.

**Beknazar-Yuzbashev, George, Rafael Jiménez-Durán, and Mateusz Stalinski**, "A model of harmful yet engaging content on social media," in "AEA Papers and Proceedings," Vol. 114 American Economic Association 2024, pp. 678–683.

__ , __ , **Jesse McCrosky, and Mateusz Stalinski**, "Toxic content and user engagement on social media: Evidence from a field experiment," CESifo Working Paper 11644, CESifo 2025. Revised May 2026.

**Bhadani, Saumya, Shun Yamaya, Alessandro Flammini, Filippo Menczer, Giovanni Luca Ciampaglia, and Brendan Nyhan**, "Political audience diversity and news reliability in algorithmic ranking," *Nature Human Behaviour*, 2022, *6* (4), 495–505.

**Bhandari, Anmol, Jaroslav Borovička, and Paul Ho**, "Survey data and subjective beliefs in business cycle models," *The Review of Economic Studies*, 2025, *92* (3), 1375–1437.

**Binetti, Alberto, Francesco Nuzzi, and Stefanie Stantcheva**, "People's understanding of inflation," *Journal of Monetary Economics*, 2024, *148*, 103652.

**Bonhomme, Stéphane and Elena Manresa**, "Grouped patterns of heterogeneity in panel data," *Econometrica*, 2015, *83* (3), 1147–1184.

**Braghieri, Luca, Ro'ee Levy, and Hannah Trachtman**, "Frictions in News Consumption: Evidence from Social Media," CESifo Working Paper 12771, CESifo 2026.

__ , **Sarah Eichmeyer, Ro'ee Levy, Markus Mobius, Jacob Steinhardt, and Ruiqi Zhong**, "Article-Level Slant and Polarization of News Consumption on Social Media," CEPR Discussion Paper 19807, Centre for Economic Policy Research 2024. Revised July 2026.

**Broockman, David E. and Joshua L. Kalla**, "Consuming Cross-Cutting Media Causes Learning and Moderates Attitudes: A Field Experiment with Fox News Viewers," *The Journal of Politics*, January 2025, *87* (1), 246–261.

**Bruine de Bruin, Wändi, Charles F. Manski, Giorgio Topa, and Wilbert van der Klaauw**, "Measuring consumer uncertainty about future inflation," *Journal of Applied Econometrics*, 2011, *26* (3), 454–478.

__ , **Wilbert van der Klaauw, Giorgio Topa, Julie S. Downs, Baruch Fischhoff, and Olivier Armantier**, "The effect of question wording on consumers' reported inflation expectations," *Journal of Economic Psychology*, 2012, *33* (4), 749–757.

**Burke, Mary A. and Ali Ozdagli**, "Household inflation expectations and consumer spending: Evidence from panel data," *The Review of Economics and Statistics*, 2023, *105* (4), 948–961.

**Bursztyn, Leonardo, Aakaash Rao, Christopher Roth, and David Yanagizawa-Drott**, "Opinions As Facts," *The Review of Economic Studies*, 2023, *90* (4), 1832–1864.

__ , **Benjamin R. Handel, Rafael Jiménez-Durán, and Christopher Roth**, "When Product Markets Become Collective Traps: The Case of Social Media," *American Economic Review*, December 2025, *115* (12), 4105–4136.

__ , **Ingar Haaland, Nicolas Röver, and Christopher Roth**, "The Social Desirability Atlas," CESifo Working Paper 11911, CESifo 2025. Forthcoming, Journal of Political Economy Microeconomics.

**Cagé, Julia, Moritz Hengel, Nicolas Hervé, and Camille Urvoy**, "Political Bias in the Media: Evidence from the Universe of French Broadcasts, 2002–2020," CESifo Working Paper 11741, CESifo 2025.

**Campante, Filipe, Ruben Durante, Felix Hagemeister, and Ananya Sen**, "GenAI Misinformation, Trust, and News Consumption: Evidence from a Field Experiment," CEPR Discussion Paper 20526, Centre for Economic Policy Research August 2025.

**Capozza, Francesco, Ingar Haaland, Christopher Roth, and Johannes Wohlfart**, "News Consumption in the Digital Age," CESifo Working Paper 12945, CESifo 2026.

**Caprini, Giulia**, "Visual Bias," Economics Series Working Paper 1016, University of Oxford, Department of Economics 2023.

**Carroll, Christopher D**, "Macroeconomic expectations of households and professional forecasters," *The Quarterly Journal of Economics*, 2003, *118* (1), 269–298.

**Carvalho, Carlos and Fernanda Nechio**, "Do people understand monetary policy?," *Journal of Monetary Economics*, 2014, *66*, 108–123.

**Chen, Canyu and Kai Shu**, "Can LLM-Generated Misinformation Be Detected?," in "The Twelfth International Conference on Learning Representations" 2024.

**Cheng, Mengjie (Magie), Elie Ofek, and Hema Yoganarasimhan**, "Balancing Engagement and Polarization: Multi-Objective Alignment of News Content Using LLMs," HBS Working Paper 25-051, Harvard Business School 2025.

**Chopra, Felix and Ingar Haaland**, "Conducting Qualitative Interviews with AI," CESifo Working Paper 10666, CESifo 2026.

**_ , _ , and Christopher Roth**, "Do People Demand Fact-checked News? Evidence from U.S. Democrats," *Journal of Public Economics*, 2022, *205*, 104549.

**_ , _ , and _** , "The Demand for News: Accuracy Concerns versus Belief Confirmation Motives," *The Economic Journal*, July 2024, *134* (661), 1806–1834.

**_ , _ , Fabian Roeben, Christopher Roth, and Vanessa Sticher**, "News Customization with AI," CESifo Working Paper 12121, CESifo 2025.

**Coibion, Olivier and Yuriy Gorodnichenko**, "What can survey forecasts tell us about information rigidities?," *Journal of Political Economy*, 2012, *120* (1), 116–159.

**_ , Dimitris Georgarakos, Yuriy Gorodnichenko, and Maarten van Rooij**, "How does consumption respond to news about inflation? Field evidence from a randomized control trial," *American Economic Journal: Macroeconomics*, 2023, *15* (3), 109–152.

**_ , Yuriy Gorodnichenko, and Michael Weber**, "Monetary policy communications and their effects on household inflation expectations," *Journal of Political Economy*, 2022, *130* (6), 1537–1584.

**—, —, and Rupal Kamdar**, "The formation of expectations, inflation, and the Phillips curve," *Journal of Economic Literature*, 2018, *56* (4), 1447–1491.

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

**D'Acunto, Francesco and Michael Weber**, "Why survey-based subjective expectations are meaningful and important," *Annual Review of Economics*, 2024, *16* (1), 329–357.

**Das, Sreyoshi, Camelia M. Kuhnen, and Stefan Nagel**, "Socioeconomic status and macroeconomic expectations," *The Review of Financial Studies*, 2020, *33* (1), 395–432.

**de Quidt, Jonathan, Johannes Haushofer, and Christopher Roth**, "Measuring and Bounding Experimenter Demand," *American Economic Review*, 2018, *108* (11), 3266–3302.

**DellaVigna, Stefano and Ethan Kaplan**, "The Fox News Effect: Media Bias and Voting," *Quarterly Journal of Economics*, 2007, *122* (3), 1187–1234.

**Demsetz, Harold and Kenneth Lehn**, "The Structure of Corporate Ownership: Causes and Consequences," *Journal of Political Economy*, 1985, *93* (6), 1155–1177.

**Djourelova, Milena**, "Persuasion through slanted language: Evidence from the media coverage of immigration," *American Economic Review*, 2023, *113* (3), 800–835.

**Dominitz, Jeff and Charles F. Manski**, "Using expectations data to study subjective income expectations," *Journal of the American Statistical Association*, 1997, *92* (439), 855–867.

**Dräger, Lena and Giang Nghiem**, "Are consumers' spending decisions in line with a Euler equation?," *The Review of Economics and Statistics*, 2021, *103* (3), 580–596.

**— and Michael J. Lamla**, "Consumers' macroeconomic expectations," *Journal of Economic Surveys*, 2024, *38* (2), 427–451.

**—, —, and Damjan Pfajfar**, "Are survey expectations theory-consistent? The role of central bank communication and news," *European Economic Review*, 2016, *85*, 84–111.

**Druckman, James N, Erik Peterson, and Rune Slothuus**, "How elite partisan polarization affects public opinion formation," *American Political Science Review*, 2013, *107* (1), 57–79.

**Duca-Radu, Ioana, Geoff Kenny, and Andreas Reuter**, "Inflation expectations, consumption and the lower bound: Micro evidence from a large multi-country survey," *Journal of Monetary Economics*, 2021, *118*, 120–134.

**Durante, Ruben and Brian Knight**, "Partisan Control, Media Bias, and Viewer Responses: Evidence from Berlusconi's Italy," *Journal of the European Economic Association*, 2012, *10* (3), 451–481.

**—, Paolo Pinotti, and Andrea Tesei**, "The Political Legacy of Entertainment TV," *American Economic Review*, July 2019, *109* (7), 2497–2530.

**Eddy, Kirsten, Katerina Eva Matsa, Michael Lipka, Naomi Forman-Katz, Christopher St. Aubin, Luxuan Wang, and Kaitlyn Radde**, "What Is News?," *Pew Research Center*, May 2025. Accessed June 4, 2025.

**Engelberg, Joseph, Charles F. Manski, and Jared Williams**, "Comparing the point predictions and subjective probability distributions of professional forecasters," *Journal of Business & Economic Statistics*, 2009, *27* (1), 30–41.

**Enke, Benjamin and Thomas Graeber**, "Cognitive uncertainty," *The Quarterly Journal of Economics*, 2023, *138* (4), 2021–2067.

**European Central Bank**, "ECB Consumer Expectations Survey: Methodological information on the survey," 2024. ECB CES website, February 2024; accessed June 8, 2026.

**Ferreira, Clodomiro and Stefano Pica**, "Households' subjective expectations: Disagreement, common drivers and reaction to monetary policy," Technical Report 2445, Banco de España 2024.

**Forman-Katz, Naomi**, "Americans are following the news less closely than they used to," *Pew Research Center*, October 24 2023.

**Fuster, Andreas, Ricardo Perez-Truglia, Mirko Wiederholt, and Basit Zafar**, "Expectations with endogenous information acquisition: An experimental investigation," *The Review of Economics and Statistics*, 2022, *104* (5), 1059–1078.

**Gentzkow, Matthew**, "Media and artificial intelligence," *Toulouse Network for Information Technology Whitepaper*, 2018.

**Gentzkow, Matthew A. and Jesse M. Shapiro**, "Media Bias and Reputation," *Journal of Political Economy*, 2006, *114* (2), 280–316.

_ **and** _ , "What Drives Media Slant? Evidence From U.S. Daily Newspapers," *Econometrica*, 2010, *78* (1), 35–71.

**Gentzkow, Matthew, Jesse M. Shapiro, and Michael Sinkinson**, "Competition and Ideological Diversity: Historical Evidence from US Newspapers," *American Economic Review*, 2014, *104* (10), 3073–3114.

**Georgarakos, Dimitris and Geoff Kenny**, "Household spending and fiscal support during the COVID-19 pandemic: Insights from a new consumer survey," *Journal of Monetary Economics*, 2022, *129*, S1–S14.

**Gerber, Alan S., Dean Karlan, and Daniel Bergan**, "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions," *American Economic Journal: Applied Economics*, April 2009, *1* (2), 35–52.

**Gillen, Ben, Erik Snowberg, and Leeat Yariv**, "Experimenting with measurement error: Techniques with applications to the Caltech cohort study," *Journal of Political Economy*, 2019, *127* (4), 1826–1863.

**González-Bailón, Sandra, David Lazer, Pablo Barberá, Meig Zhang, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Deen Freelon, Matthew Gentzkow, Andrew M Guess et al.**, "Asymmetric ideological segregation in exposure to political news on Facebook," *Science*, 2023, *381* (6656), 392–398.

**Gottfried, Jeffrey**, "Americans' news fatigue isn't going away – about two-thirds still feel worn out," *Pew Research Center*, 2020.

**Graeber, Thomas, Christopher Roth, and Constantin Schesch**, "Explanations," CESifo Working Paper 11131, CESifo 2024.

**Green, Jon, Stefan McCabe, Sarah Shugars, Hanyu Chwe, Luke Horgan, Shuyang Cao, and David Lazer**, "Curation bubbles," *American Political Science Review*, 2025, *119* (4), 1704–1722.

**Groseclose, Tim and Jeffrey Milyo**, "A measure of media bias," *Quarterly Journal of Economics*, 2005, *120* (4), 1191–1237.

**Guess, Andrew M.**, "(Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets," *American Journal of Political Science*, 2021, *65* (4), 1007–1022.

**_ , Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow, Sandra González-Bailón, Edward Kennedy, Young Mie Kim, David Lazer, Devra Moehler, Brendan Nyhan, Carlos Velasco Rivera, Jaime Settle, Daniel Robert Thomas, Emily Thorson, Rebekah Tromble, Arjun Wilkins, Magdalena Wojcieszak, Beixian Xiong, Chad Kiewiet de Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud, and Joshua A. Tucker**, "How do social media feed algorithms affect attitudes and behavior in an election campaign?," *Science*, July 2023, *381* (6656), 398–404. Publisher: American Association for the Advancement of Science.

**_ , _ , _ , _ , _ , _ , _ , _ , _ , _ , _ , _ , _ , _ , _ , _ , _ , _ , _ , _ , _ , _ , _ , _ , Chad Kiewiet De Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud, and Joshua A. Tucker**, "Reshares on social media amplify political news but do not detectably affect beliefs or opinions," *Science*, July 2023, *381* (6656), 404–408.

**_ , Pablo Barberá, Simon Munzert, and JungHwan Yang**, "The consequences of online partisan media," *Proceedings of the National Academy of Sciences*, April 2021, *118* (14), e2013464118. Publisher: Proceedings of the National Academy of Sciences.

**Haaland, Ingar K, Christopher Roth, Stefanie Stantcheva, and Johannes Wohlfart**, "Understanding Economic Behavior Using Open-Ended Survey Data," *Journal of Economic Literature*, December 2025, *63* (4), 1244–1280.

**Habibi, Mahyar, Dirk Hovy, and Carlo Schwarz**, "The Content Moderator's Dilemma: Removal of Toxic Content and Distortions to Online Discourse," CEPR Discussion Paper 21257, Centre for Economic Policy Research 2026.

**Hurd, Michael D.**, "Subjective probabilities in household surveys," *Annual Review of Economics*, 2009, *1* (1), 543–562.

**Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J. Westwood**, "The Origins and Consequences of Affective Polarization in the United States," *Annual Review of Political Science*, 2019, *22* (1), 129–146.

**Jiménez-Durán, Rafael**, "The economics of content moderation: Theory and experimental evidence from hate speech on Twitter," Stigler Center Working Paper 324, George J. Stigler Center for the Study of the Economy and the State 2023.

**_ , Karsten Müller, and Carlo Schwarz**, "The effect of content moderation on online and offline hate: Evidence from Germany's NetzDG," CAGE Working Paper 701, Centre for Competitive Advantage in the Global Economy 2024.

**Jo, Donghee**, "Better the Devil You Know: Selective Exposure Alleviates Polarization in an Online Field Experiment," Technical Report, Working Paper 2019.

**Kalra, Aarushi**, "Hate in the Time of Algorithms: Evidence on Online Behavior from a Large-Scale Experiment," 2025. Working paper, Brown University.

**Kamdar, Rupal and Walker Ray**, "Attention-driven sentiment and the business cycle," Technical Report 18984, Centre for Economic Policy Research 2024.

**Kim, Gwangmin and Carola Binder**, "Learning-through-survey in inflation expectations," *American Economic Journal: Macroeconomics*, 2023, *15* (2), 254–278.

**Kleinberg, Jon, Sendhil Mullainathan, and Manish Raghavan**, "The challenge of understanding what users want: Inconsistent preferences and engagement optimization," *Management Science*, 2024, *70* (9), 6336–6355.

**Kuchler, Theresa and Basit Zafar**, "Personal experiences and expectations about aggregate outcomes," *The Journal of Finance*, 2019, *74* (5), 2491–2542.

**Leiser, David and Ronen Aroch**, "Lay understanding of macroeconomic causation: The good-begets-good heuristic," *Applied Psychology*, 2009, *58* (3), 370–384.

**_ and Zeev Krill**, "How laypeople understand the economy," in Rob Ranyard, ed., *Economic Psychology*, John Wiley & Sons, 2017, pp. 139–154.

**Levy, Ro'ee**, "Social Media, News Consumption, and Polarization: Evidence from a Field Experiment," *American Economic Review*, March 2021, *111* (3), 831–70.

**Link, Sebastian, Andreas Peichl, Christopher Roth, and Johannes Wohlfart**, "Information frictions among firms and households," *Journal of Monetary Economics*, 2023, *135*, 99–115.

**Liu, Naijia, Xinlan Emily Hu, Yasemin Savas, Matthew A. Baum, Adam J. Berinsky, Allison J. B. Chaney, Christopher Lucas, Rei Mariman, Justin de Benedictis-Kessner, Andrew M. Guess, Dean Knox, and Brandon M. Stewart**, "Short-term exposure to filter-bubble recommendation systems has limited polarization effects: Naturalistic experiments on YouTube," *Proceedings of the National Academy of Sciences*, February 2025, *122* (8), e2318127122. Publisher: Proceedings of the National Academy of Sciences.

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

**Mankiw, N. Gregory, Ricardo Reis, and Justin Wolfers**, "Disagreement about inflation expectations," *NBER Macroeconomics Annual*, 2003, *18*, 209–248.

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

_, "Survey Measurement of Probabilistic Macroeconomic Expectations: Progress and Promise," *NBER Macroeconomics Annual*, 2018, *32* (1), 411–471.

_ **and Francesca Molinari**, "Rounding probabilistic expectations in surveys," *Journal of Business & Economic Statistics*, 2010, *28* (2), 219–231.

**Martin, Gregory J and Ali Yurukoglu**, "Bias in cable news: Persuasion and polarization," *American Economic Review*, 2017, *107* (9), 2565–2599.

**Min, Sewon, Kalpesh Krishna, Xinxi Lyu, Mike Lewis, Wen tau Yih, Pang Wei Koh, Mohit Iyyer, Luke Zettlemoyer, and Hannaneh Hajishirzi**, "FActScore: Fine-Grained Atomic Evaluation of Factual Precision in Long Form Text Generation," in "Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing" Association for Computational Linguistics Singapore December 2023, pp. 12076–12100.

**Mitchell, Amy, Jeffrey Gottfried, Michael Barthel, and Nami Sumida**, "Distinguishing Between Factual and Opinion Statements in the News," Technical Report, Pew Research Center June 2018.

_ **, Katie Simmons, Katerina Eva Matsa, and Laura Silver**, "Publics Globally Want Unbiased News Coverage, but Are Divided on Whether Their News Media Deliver," https://www.pewresearch.org/journalism/2018/01/11/publics-globally-want-unbiased-news-coverage-but-are-divided-on-whether-their-news-media-deliver/ 2018. Pew Research Center.

**Mitra, Anushka and Aditi Singh**, "What determines household expectations?," Technical Report 1412, Board of Governors of the Federal Reserve System 2025.

**Mullainathan, Sendhil and Andrei Shleifer**, "The Market for News," *American Economic Review*, 2005, *95* (4), 1031–1053.

**Ni, Jingwei, Minjing Shi, Dominik Stammbach, Mrinmaya Sachan, Elliott Ash, and Markus Leippold**, "AFaCTA: Assisting the Annotation of Factual Claim Detection with Reliable LLM Annotators," in "Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)" Association for Computational Linguistics Bangkok, Thailand August 2024, pp. 1890–1912.

**Noy, Shakked and Aakaash Rao**, "The Business of the Culture War," December 2025. Job Market Paper.

**Nyhan, Brendan and Jason Reifler**, "When corrections fail: The persistence of political misperceptions," *Political Behavior*, 2010, *32* (2), 303–330.

— , **Jaime Settle, Emily Thorson, Magdalena Wojcieszak, Pablo Barberá, Annie Y. Chen, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow, Sandra González-Bailón, Andrew M. Guess, Edward Kennedy, Young Mie Kim, David Lazer, Neil Malhotra, Devra Moehler, Jennifer Pan, Daniel Robert Thomas, Rebekah Tromble, Carlos Velasco Rivera, Arjun Wilkins, Beixian Xiong, Chad Kiewiet de Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud, and Joshua A. Tucker**, "Like-minded sources on Facebook are prevalent but not polarizing," *Nature*, August 2023, *620* (7972), 137–144. Publisher: Nature Publishing Group.

**Peer, Eyal, David Rothschild, Ariel Gordon, Zak Evernden, and Ekaterina Damer**, "Data quality of platforms and panels for online behavioral research," *Behavior Research Methods*, 2022, *54* (4), 1643–1662.

**Prat, Andrea and David Strömberg**, "The Political Economy of Mass Media," in Richard Blundell, Whitney Newey, and Torsten Persson, eds., *Advances in Economics and Econometrics: Tenth World Congress*, Vol. 2, Cambridge University Press, 2013, pp. 135–187.

**Prior, Markus**, *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*, Cambridge University Press, 2007.

**Qin, Bei, David Strömberg, and Yanhui Wu**, "Media Bias in China," *American Economic Review*, 2018, *108* (9), 2442–2476.

**Reimers, Nils and Iryna Gurevych**, "Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks," in "Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing" Association for Computational Linguistics Hong Kong, China November 2019, pp. 3982–3992.

**Reuters**, "Digital News Report 2023: Executive Summary," Technical Report, Reuters Institute for the Study of Journalism 2023. Accessed: December 1, 2024.

**Rilla, Raluca, Tobias Werner, Hiromu Yakura, Iyad Rahwan, and Anne-Marie Nussberger**, "Recognising and Mitigating LLM Pollution in Online Behavioural Research," *Nature Communications*, June 2026, *17* (1), 5578.

**Robertson, Claire E, Nicolas Pröllochs, Kaoru Schwarzenegger, Philip Pärnamets, Jay J Van Bavel, and Stefan Feuerriegel**, "Negativity drives online news consumption," *Nature Human Behaviour*, 2023, *7* (5), 812–822.

**Roth, Christopher and Johannes Wohlfart**, "How do expectations about the macroeconomy affect personal expectations and behavior?," *The Review of Economics and Statistics*, 2020, *102* (4), 731–748.

**Sunstein, Cass R**, # *Republic: Divided Democracy in the Age of Social Media*, Princeton University Press, 2018.

**Thaler, Michael**, "The fake news effect: Experimentally identifying motivated reasoning using trust in news," *American Economic Journal: Microeconomics*, 2024, *16* (2), 1–38.

**Weber, Michael, Francesco D'Acunto, Yuriy Gorodnichenko, and Olivier Coibion**, "The subjective inflation expectations of households and firms: Measurement, determinants, and implications," *Journal of Economic Perspectives*, 2022, *36* (3), 157–184.

**Wei, Jerry, Chengrun Yang, Xinying Song, Yifeng Lu, Nathan Hu, Jie Huang, Dustin Tran, Daiyi Peng, Ruibo Liu, Da Huang, Cosmo Du, and Quoc V. Le**, "Long-Form Factuality in Large Language Models," in "Advances in Neural Information Processing Systems 37," Vol. 37 2024, pp. 80756–80827.

**Zhou, Jiawei, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury**, "Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions," in "Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems" Association for Computing Machinery 2023, pp. 1–20.

**Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov**, "Political effects of the internet and social media," *Annual Review of Economics*, 2020, *12* (1), 415–438.

# Appendix A: News Customization with AI

## A.1 Summary of Data Collections

Table A.1: Overview of Data Collections

| Data collection | Sample | Treatment arms | Main outcomes |
|---|---|---|---|
| **Human Validation Wave 1** (December 2, 2024) | Prolific ($n$ = 500) | No treatments | Rating of articles in terms of complexity, entertainment, political leaning and sentiment |
| **Human Validation Wave 2** (August 19-21, 2025) | Prolific ($n$ = 669) | No treatments | Rating of articles in terms of facts, complexity, entertainment, political leaning, sentiment |
| **Latent Demand Experiment** (December 13 and 14 2024) | Prolific ($n = 5,005$) | No Customization (Regular), Customization with political rewrites, Customization with *Facts Only* rewrites | Download and log in to the app |
| **Usage Incentive Experiment Wave 1** (October and November, 2024) | Prolific ($n = 1,439$) | Customization with political rewrites, No Customization (Regular, Less Complex, Entertaining, Liberal, Conservative, More Positive, More Negative) | Time spent on app (in different versions), knowledge acquisition, political polarization |
| **Usage Incentive Experiment Wave 2** (January and February, 2025) | Prolific ($n = 2,039$) | Customization with political rewrites, Customization with Facts Only rewrites, No Customization (Regular, Less Complex, Entertaining, Liberal, Conservative, More Positive, More Negative, Facts Only) | Time spent on app (in different versions), knowledge acquisition, political polarization |

*Note:* Table A.1 displays the overview of the data collections. All data collections were pre-registered on the AEA RCT registry (#14561). The experimental instructions for each collection can be found on the following link: https://raw.githubusercontent.com/cproth/papers/master/tmt_instructions_appendix.pdf.

## A.2 Validation of News Event Consistency

We validate that rewrites preserve the underlying news event using a custom three-step semantic fact-level alignment procedure applied to the full corpus of 59,092 article versions across both waves of the experiment. For all LLM usage in this pipeline, we ensure reproducible outputs by disabling sampling (temperature = 0, top_p = 1) and setting both frequency and presence penalties to zero.[1]

**Extraction** We define a factual claim as a single, objectively verifiable statement, following Ni et al. (2024). This definition is consistent with recent advances in automatic factuality evaluation, which assess content at the atomic level (Min et al., 2023; Wei et al., 2024). Facts were extracted sentence by sentence from each article version using GPT-4.1 (nano). To ensure high-quality extraction for every sentence, we provided the model with the three most relevant examples from a hand-annotated library of 96 sentences, following the approach of Min et al. (2023). Relevance was determined using a standard text-matching algorithm (BM25), which ensures the model receives highly contextual demonstrations. This retrieval-augmented generation (RAG) setup grounds the model in closely related examples, reduces variance across domains, and achieves high extraction accuracy. Further, outputs were parsed and schema-validated as well as near-duplicates removed. As a final quality filter, we retained only those extractions that formed complete, standalone statements (i.e., finite clauses). This step ensures that every fact is a grammatically whole proposition that can be independently verified, discarding sentence fragments or incomplete thoughts. Lastly, numeric metadata were preserved for later validation checks. An example extraction for a *Regular* version appears in Table A.2.

**Classification** The extracted facts are classified into core or supplementary facts using GPT-4.1. Core facts capture the "who did what and why" of the main event as well as its primary intent and consist of the set sufficient to form a 1-2 sentence, faithful account of the article. Supplementary facts cover background, context, reactions, sourcing/meta information, or other details not essential for understanding the main event. To improve consistency of the fact extraction procedure, the model first resolved pronouns using the article text before classification, and its output was constrained to a structured JSON schema, ensuring every fact was assigned to exactly one category. Two safeguards were applied: ambiguous or uncertain cases were conservatively assigned to supplementary, and duplicate facts were reduced to the single strongest version in core. Table A.2 illustrates the resulting split. Outputs were also periodically inspected during development to ensure rule adherence and to refine prompts.

**Consistency** We then tested whether core facts were consistent across article versions. For each comparison and direction, we embedded all facts using the sentence-transformer model `all-mpnet-base-v2` (unit-normalized) and computed cosine similarities. A source core fact is counted as *covered* if *any* target-side fact (core $\cup$ supplementary) attains similarity $\geq \tau$ with the source ($\tau$=0.66). This threshold was selected by maximizing F1 on 400 hand-labeled fact pairs,

[1] *Temperature* rescales token logits, where lower values reduce stochasticity. *Top_p* restricts draws to the smallest set of tokens whose cumulative probability exceeds $p$. Setting temperature = 0 and top_p = 1 yields greedy decoding in practice. *Frequency* and *presence* penalties down-weight previously generated tokens to discourage repetition; both were set to zero to avoid altering base probabilities.

Table A.2: Example of Fact Extraction and Classification

| |
|---|
| **Core Facts** |
| 1. Republican lawmakers are voicing concerns about President Trump's government reforms. |
| 2. Specific measures are drawing pushback from Republican lawmakers. |
| 3. The Department of Government Efficiency (DOGE) is led by Elon Musk. |
| 4. The government reforms are under the Department of Government Efficiency (DOGE). |
| 5. The pushback is due to potential negative impacts on home-state industries, jobs, and programs. |
| **Supplementary Facts** |
| 6. Constituents in the Research Triangle expressed concerns about potential damage to scientific progress. |
| 7. Critics argued that these measures exceed presidential authority. |
| 8. House Speaker Mike Johnson praised the administration's efforts as necessary reforms. |
| 9. Idaho Representative Mike Simpson highlighted concerns about reduced hiring at national parks ahead of the summer tourism season. |
| 10. Other Republicans, including Senator Susan Collins, detailed how funding caps could undermine critical biomedical research. |
| 11. Representative Carlos Gimenez is advocating for Venezuelans under Temporary Protected Status in Florida. |
| 12. Representative Carlos Gimenez is urging the administration to evaluate individual cases to prevent deportations. |
| 13. Senator Jerry Moran is working on legislation to transfer food aid management to the Department of Agriculture. |
| 14. Senator Jerry Moran of Kansas raised alarms about disruptions to the U.S. Agency for International Development's food aid. |
| 15. Senator Katie Britt warned that these reductions could harm essential research at institutions like the University of Alabama. |
| 16. Senator Katie Britt of Alabama has criticized cuts to National Institutes of Health (NIH) grants. |
| 17. Senator Ted Budd is a senator representing North Carolina. |
| 18. Senator Ted Budd reported hearing from constituents in the Research Triangle. |
| 19. Senate Minority Leader Chuck Schumer accused the administration of bypassing congressional debate. |
| 20. Some Republican leaders have defended the administration's efforts. |
| 21. The disruptions to USAID's food aid risk spoiling American-grown produce intended for global distribution. |
| 22. The executive actions by Musk's DOGE team have sparked legal challenges across the country. |

*Note:* This table illustrates fact extraction and classification for the *Regular* article version of the example article.

following Reimers and Gurevych (2019). This procedure yielded three sets per comparison: matched pairs, unmatched facts from the *Regular* version, and unmatched facts from the respective rewritten version. Table A.3 visualizes the embedding step on our running example: five of six source core facts meet the threshold and are matched; the final row (similarity = 0.59) falls below $\tau$ and remains unmatched under embeddings alone.

As a final validation step to reduce false negatives, each unmatched source core fact is presented together with the full target article to a state-of-the-art LLM (GPT-5 at the time of writing) under a strict JSON schema to assess whether the fact is nonetheless substantively covered. In the example from Table A.3, the LLM adjudication marks the last unmatched pair as *covered* with the justification: "Article explicitly says concerns stem from impacts on home-state industries, jobs, and programs."

As an additional diagnostic, we directly compared the number metadata preserved in the extraction step across article versions. This check focused on numerical values (e.g., dates, statistics, percentages). Consistency across these dimensions provides assurance that rewrites did not alter, add or remove quantitative information, complementing the semantic alignment procedure.

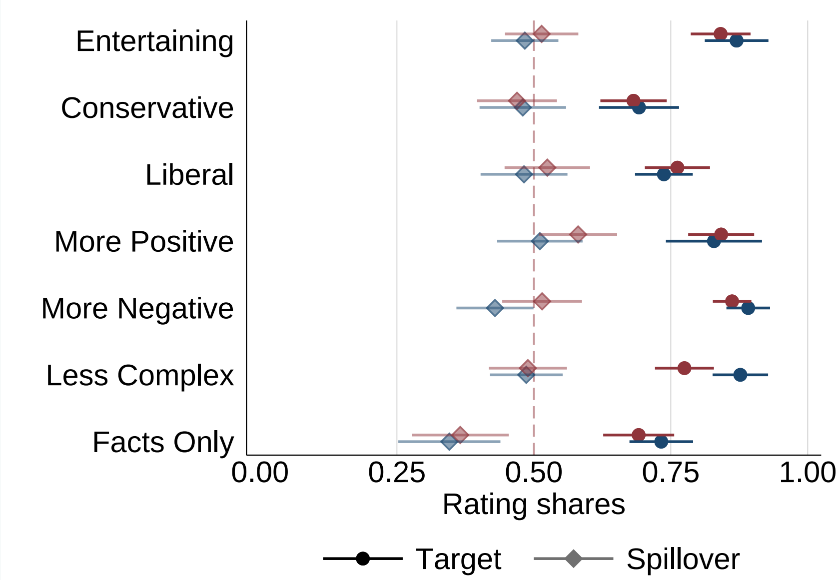
Table A.3: Example Facts Alignment

| **Facts from *Regular* version** | **Facts from *Less Complex* version** | **Similarity** | **Match** |
|---|---|---|---|
| The Department of Government Efficiency (DOGE) is led by Elon Musk. | The Department of Government Efficiency is led by billionaire Elon Musk. | 0.92 | ✓ |
| Republican lawmakers are voicing concerns about President Trump's government reforms. | Some Republican lawmakers are worried about President Trump's major changes to the government. | 0.88 | ✓ |
| *The executive actions by Musk's DOGE team have sparked legal challenges across the country.* | The actions taken by Elon Musk's Department of Government Efficiency team led to legal challenges across the country. | 0.86 | ✓ |
| The government reforms are under the Department of Government Efficiency (DOGE). | The changes are being managed by the Department of Government Efficiency. | 0.76 | ✓ |
| Specific measures are drawing pushback from Republican lawmakers. | *Some Republican lawmakers are worried about how certain changes might hurt people and important programs.* | 0.68 | ✓ |
| The pushback is due to potential negative impacts on home-state industries, jobs, and programs. | *Some Republican lawmakers are concerned that President Trump's major changes to the government could hurt industries in their states.* | 0.59 | × |

*Note:* This table shows the matching of core facts for the *Regular* and *Less Complex* version of the example article. Facts in italics represent supplementary facts from the respective article version. If a pair of facts has cosine similarity $\geq 0.66$, it is considered a match.

## A.3 Human Validation Study

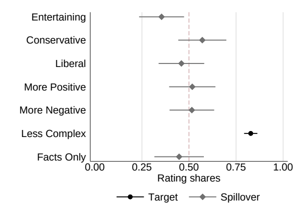
Figure A.1: Human Validation of Article Rewrites by Party Affiliation



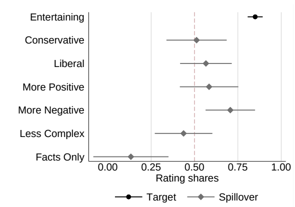

*Note:* This figure shows the share of respondents rating the rewritten article version as more entertaining, conservative, liberal, positive, negative, or less complex than the regular version in a blinded binary comparison, by party affiliation. Red denotes ratings by Republicans, while blue denotes ratings by Democrats. Dots show the share correctly identifying the version varied along the target dimension (e.g., *Entertaining* rewrites rated as more entertaining). Squares show spillovers (e.g., non-*Entertaining* rewrites rated as more entertaining). Error bars represent 95% confidence intervals.

Figure A.2: Human Validation by Rewrite Dimension

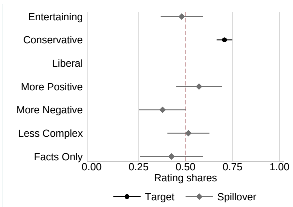
(a) Less Complex



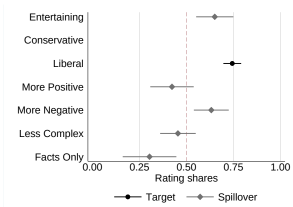

(b) Entertaining



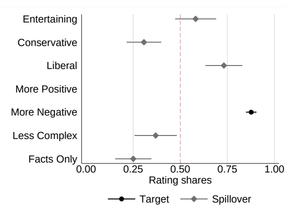

(c) Conservative



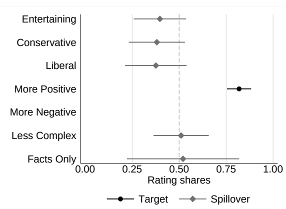

(d) Liberal



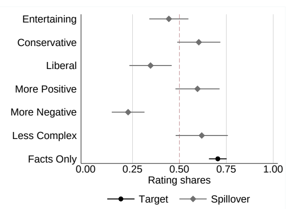

(e) More Negative




(f) More Positive



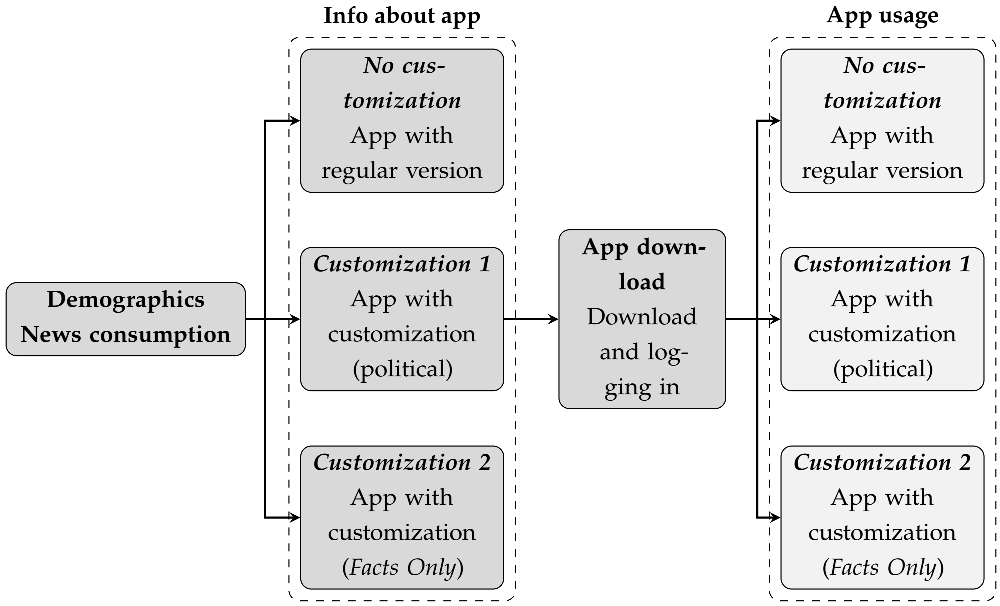

(g) Facts Only




*Note:* This figure shows the share of respondents rating the rewritten article version as more entertaining, conservative, liberal, positive, negative, or less complex compared to the regular version in a blinded binary comparison separately by rewrite version. Each panel reports ratings across all dimensions for one version. Black dots indicate correctly identifying the varied dimension (e.g., *entertaining* rewrites rated as more entertaining). Gray squares indicate spillovers (e.g., *Less Complex* rewrites rated as more entertaining). Error bars represent 95% confidence intervals.

## A.4 Demand Experiment

Figure A.3: Overview of the Demand Experiment




*Note:* This figure presents an overview of the experiment to study the latent demand for customization. The experiment starts with a survey eliciting demographics and baseline news consumption. Participants are then randomized into three conditions and provided information about the app (with or without customization, depending on the condition). Next, they can download the app. Finally, we observe their app usage.

Table A.4: Summary Statistics for the Demand Experiment

| | Mean | SD | Min | Max |
|---|---|---|---|---|
| **Demographics** | | | | |
| Age | 40.04975 | 13.38931 | 18 | 83 |
| Female | .5428571 | .4982097 | 0 | 1 |
| White | .6973027 | .4594712 | 0 | 1 |
| At least some college | .8587413 | .3483231 | 0 | 1 |
| **Political leaning** | | | | |
| Democrat | .4207792 | .4937335 | 0 | 1 |
| Independent | .3302697 | .4703571 | 0 | 1 |
| Republican | .248951 | .4324486 | 0 | 1 |
| **News consumption** | | | | |
| News primarily online | .8435564 | .3633116 | 0 | 1 |
| Somewhat frequent news | .5524476 | .4972913 | 0 | 1 |
| News apps | 1.417582 | 1.588154 | 0 | 6 |
| Observations | 5005 | | | |

*Note:* This table presents summary statistics for the demand experiment. “Age” is the respondent’s age in years. “Female”, “White”, and “At least some college” are binary indicators for gender, race, and highest level of education, respectively. Party affiliation is captured by the mutually exclusive binary indicators “Democrat”, “Independent”, and “Republican”. “News primarily online” is an indicator that equals one if the respondent consumes news primarily online; “Somewhat frequent news” equals one if the respondent consumes news at least somewhat frequently. “News apps” is the number of news apps on a respondent’s phone.

Table A.5: Balance Table for the Demand Experiment

| | (1)<br>No customization | (2)<br>Customization (political) | (3)<br>Customization (Facts Only) | (4)<br>p-val (2) vs. (1) | (5)<br>p-val (3) vs. (1) |
|---|---|---|---|---|---|
| **Demographics** | | | | | |
| Age | 39.707 | 40.166 | 40.276 | 0.323 | 0.222 |
| | (13.519) | (13.297) | (13.352) | | |
| Female | 0.540 | 0.552 | 0.536 | 0.479 | 0.810 |
| | (0.499) | (0.497) | (0.499) | | |
| White | 0.699 | 0.686 | 0.707 | 0.408 | 0.611 |
| | (0.459) | (0.464) | (0.455) | | |
| Some college | 0.873 | 0.852 | 0.851 | 0.082 | 0.069 |
| | (0.333) | (0.355) | (0.356) | | |
| **Political leaning** | | | | | |
| Democrat | 0.415 | 0.423 | 0.424 | 0.655 | 0.627 |
| | (0.493) | (0.494) | (0.494) | | |
| Independent | 0.329 | 0.336 | 0.326 | 0.686 | 0.844 |
| | (0.470) | (0.472) | (0.469) | | |
| Republican | 0.255 | 0.241 | 0.250 | 0.342 | 0.735 |
| | (0.436) | (0.428) | (0.433) | | |
| **News consumption** | | | | | |
| News primarily online | 0.853 | 0.837 | 0.842 | 0.205 | 0.379 |
| | (0.355) | (0.370) | (0.365) | | |
| Somewhat frequent news | 0.550 | 0.545 | 0.562 | 0.791 | 0.462 |
| | (0.498) | (0.498) | (0.496) | | |
| News apps | 1.405 | 1.412 | 1.436 | 0.889 | 0.572 |
| | (1.578) | (1.586) | (1.601) | | |
| Observations | 1668 | 1671 | 1666 | | |

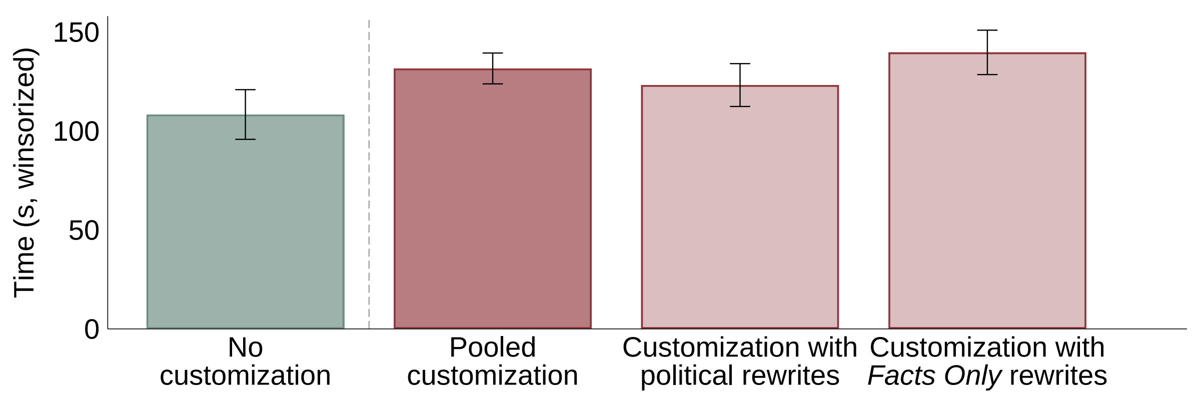
*Note:* This table reports summary statistics for the demand experiment by customization condition. “Age” is in years. “Female”, “White”, and “At least some college” are binary indicators for gender, race, and highest level of education, respectively. Party affiliation is captured by mutually exclusive indicators for “Democrat”, “Independent”, and “Republican”. “News primarily online” equals one if the respondent consumes news primarily online; “Somewhat frequent news” equals one if they consumes news at least somewhat frequently. “News apps” counts the number of news apps on a respondent’s phone. Columns (1)-(3) show means with standard deviations in parentheses for the no-customization condition and the two customization conditions. Columns (4) and (5) report p-values testing equality of no-customization and each customization conditions,.

Table A.6: App Downloads by Customization Condition

| | (1) Downloaded | (2) Downloaded | (3) Downloaded | (4) Downloaded |
|---|---|---|---|---|
| Customization | 0.0272*** | 0.0259*** | | |
| | (0.00801) | (0.00798) | | |
| Customization (political) | | | 0.0256*** | 0.0246*** |
| | | | (0.00944) | (0.00942) |
| Customization (Facts Only) | | | 0.0289*** | 0.0272*** |
| | | | (0.00954) | (0.00951) |
| Mean with no customization | 0.0683 | 0.0683 | 0.0683 | 0.0683 |
| P-value: pol=Facts Only | | | 0.747 | 0.798 |
| N | 5005 | 5005 | 5005 | 5005 |
| Controls | No | Yes | No | Yes |

*Note:* This table presents regression coefficients for the effect of customization on app downloads. "Customization" equals one for respondents in any customization condition (pooled). Columns (1)-(2) show pooled effects. Columns (3)-(4) separate the two customization conditions: political and *Facts Only* rewrites. Columns (2) and (4) control for demographics (age, gender, race, income, some college, full-time employed, party affiliation) and news consumption (primarily consuming news online, number of news apps on the phone). Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

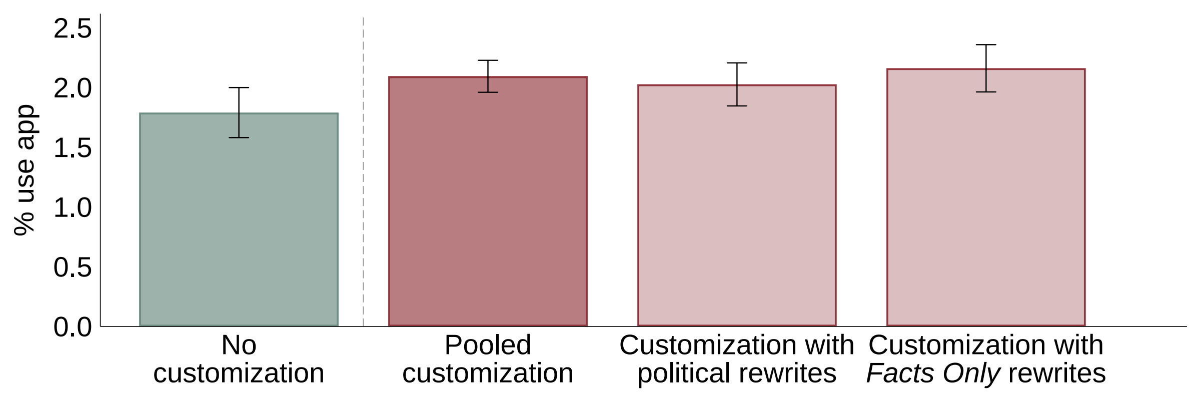
Figure A.4: App Usage Time on Day 1 by Customization Condition



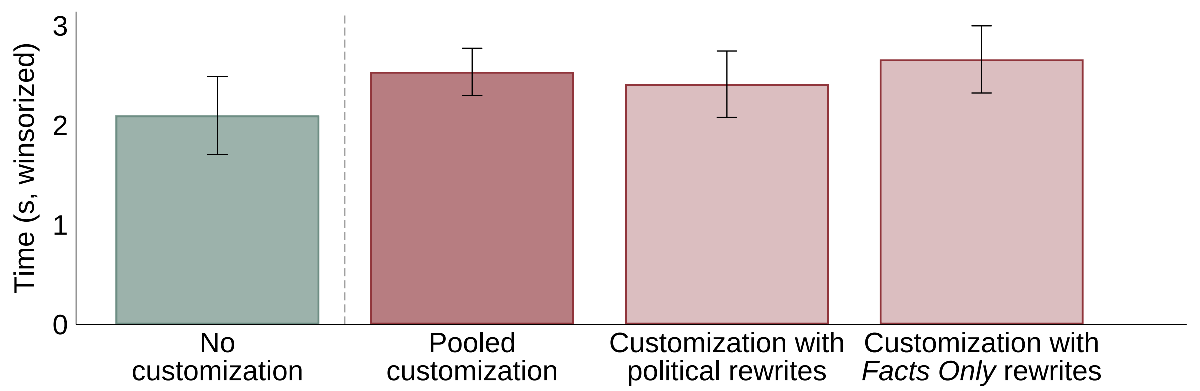

*Note:* This figure shows app usage time on day 1 by customization, conditional on app download. Time is in seconds and winsorized at the 98th percentile. The first bar shows the no customization condition, the second bar shows both customization conditions pooled, the last two bars show customization conditions separately (political and *Facts Only*). Error bars represent standard errors of the mean.

Figure A.5: App Usage Frequency and Time by Customization Condition

(a) App Usage Frequency




(b) Average Time Spent

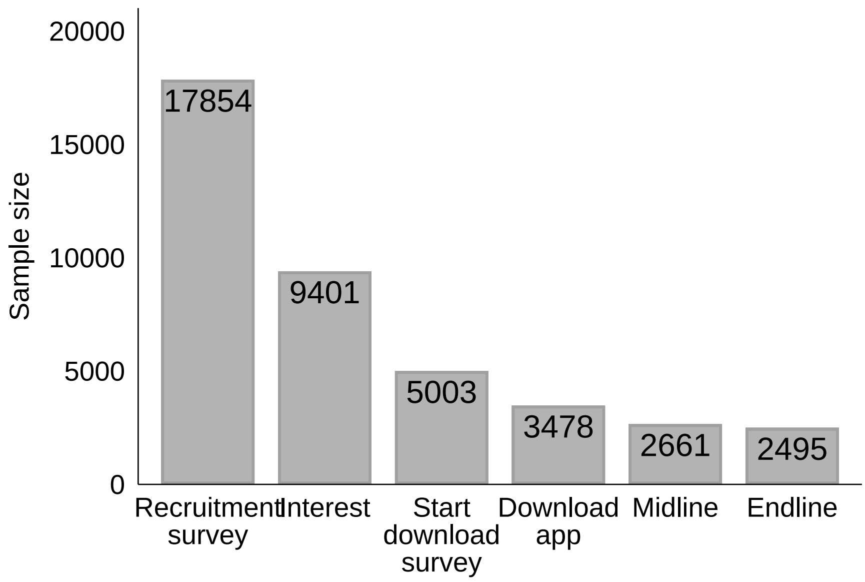



*Note:* This figure shows daily app usage for the first seven days after the download survey by customization condition. Panel (a) shows the average daily share of users. Panel (b) shows average daily usage time in seconds and winsorized at the 98th percentile. In each panel, the first bar shows no customization, the second bar pools both customization conditions, and the last two bars show political and *Facts Only* separately. Error bars represent standard errors of the mean.

## A.5 Usage Incentive Experiment

### A.5.1 Sample

Figure A.6: Sample Size in the Usage Incentive Experiment




*Note:* This figure shows the sample sizes at the different stages of the usage incentive experiment. The bars represent the number of participants who completed the recruitment survey, expressed interest in early access to the app, started the download survey, downloaded the app, and completed the midline and endline surveys.

Table A.7: Treatment Arms in the Usage Incentive Experiment

| | Total | Wave 1 | Wave 2 |
|---|---|---|---|
| Customization (political) | 747 | 430 | 317 |
| Customization (Facts Only) | 718 | | 718 |
| Regular | 261 | 140 | 121 |
| Liberal | 237 | 135 | 102 |
| Conservative | 254 | 151 | 103 |
| Less Complex | 238 | 131 | 107 |
| More Negative | 264 | 158 | 106 |
| More Positive | 254 | 146 | 108 |
| Entertaining | 263 | 148 | 115 |
| Facts Only | 242 | | 242 |
| *N* | 3478 | 1439 | 2039 |

*Note:* This table reports the number of respondents in each treatment arm, both in total (Column 1) and separately for Wave 1 and Wave 2 (Columns 2 and 3, respectively). The first two rows show to the two customization conditions. The remaining rows show to the control conditions, where respondents were shown one randomly assigned rewrite version. The *Facts Only* conditions were only introduced in Wave 2 and therefore over-sampled to achieve comparable final sample sizes. Target sample sizes were 750 participants per customization condition and 250 per control condition. While exact targets could not be achieved due to technical constraints, the final sample sizes closely approximate the targets.

Table A.8: Attrition in the Usage Incentive Experiment

| | (1) Midline | (2) Endline |
|---|---|---|
| Customization (political) | -0.00834 | -0.00873 |
| | (0.0177) | (0.0190) |
| Customization (Facts Only) | 0.00343 | 0.00306 |
| | (0.0202) | (0.0211) |
| Constant | 0.766*** | 0.719*** |
| | (0.00939) | (0.0100) |
| Download day FE | Yes | Yes |
| N | 3478 | 3478 |

*Note:* This table reports regression estimates of attrition by treatment arm in the usage incentive experiment. The dependent variable is a binary indicator for completing the midline survey (Column 1) or the endline survey (Column 2). Regressions include fixed effects for the day on which the app download survey was completed. Coefficients for both customization conditions are close to zero, indicating no evidence of differential attrition across groups. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9: Summary Statistics for the Usage Incentive Experiment

| | Mean | SD | Min | Max |
|---|---|---|---|---|
| **Demographics** | | | | |
| Age | 37.71794 | 12.14871 | 18 | 78 |
| Female | .5644048 | .495906 | 0 | 1 |
| White | .6912018 | .4620641 | 0 | 1 |
| At least some college | .8780909 | .3272279 | 0 | 1 |
| **Political leaning** | | | | |
| Democrat | .4278321 | .4948355 | 0 | 1 |
| Independent | .3323749 | .4711323 | 0 | 1 |
| Republican | .239793 | .4270184 | 0 | 1 |
| **News consumption** | | | | |
| News primarily online | .9045428 | .2938876 | 0 | 1 |
| Somewhat frequent news | .5802185 | .493594 | 0 | 1 |
| News apps | 1.780046 | 1.611072 | 0 | 6 |
| Observations | 3478 | | | |

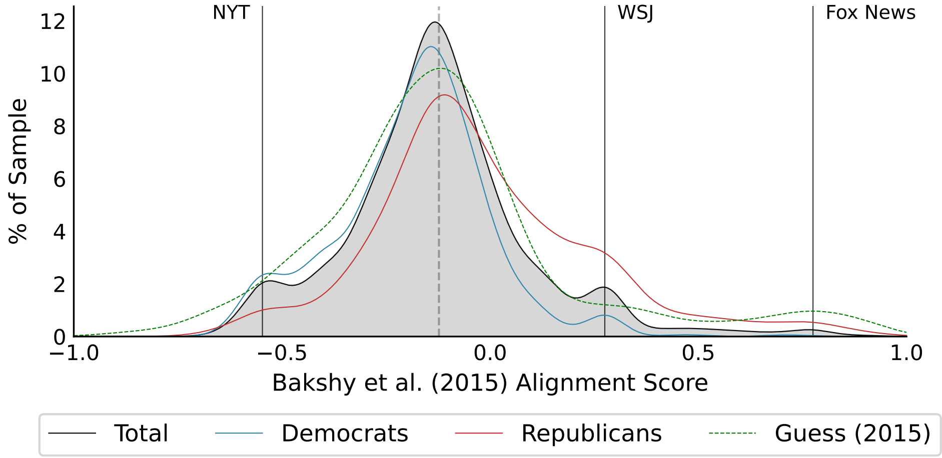
*Note:* This table presents summary statistics for the usage incentive experiment. “Age” is the respondent’s age in years. “Female”, “White”, and “At least some college” are binary indicators for gender, race, and highest level of education, respectively. Party affiliation is captured by the mutually exclusive binary indicators “Democrat”, “Independent”, and “Republican”. “News primarily online” is an indicator that equals one if the respondent consumes news primarily online; “Somewhat frequent news” equals one if the respondent consumes news at least somewhat frequently. “News apps” is the number of news apps on a respondent’s phone.

Table A.10: Balance Table for the Usage Incentive Experiment

| | (1)<br>No customization | (2)<br>Customization (political) | (3)<br>Customization (Facts Only) | (4)<br>p-val (2) vs. (1) | (5)<br>p-val (3) vs. (1) |
|---|---|---|---|---|---|
| **Demographics** | | | | | |
| Age | 37.456 | 37.901 | 38.262 | 0.331 | 0.819 |
| | (12.072) | (12.460) | (12.031) | | |
| Female | 0.567 | 0.576 | 0.546 | 0.907 | 0.688 |
| | (0.496) | (0.495) | (0.498) | | |
| White | 0.697 | 0.680 | 0.685 | 0.420 | 0.406 |
| | (0.459) | (0.467) | (0.465) | | |
| At least some college | 0.880 | 0.901 | 0.850 | 0.167 | 0.524 |
| | (0.325) | (0.299) | (0.358) | | |
| **Political leaning** | | | | | |
| Democrat | 0.443 | 0.411 | 0.404 | 0.076 | 0.519 |
| | (0.497) | (0.492) | (0.491) | | |
| Independent | 0.328 | 0.336 | 0.340 | 0.541 | 0.620 |
| | (0.470) | (0.473) | (0.474) | | |
| Republican | 0.229 | 0.253 | 0.256 | 0.176 | 0.201 |
| | (0.420) | (0.435) | (0.437) | | |
| **News consumption** | | | | | |
| News primarily online | 0.907 | 0.900 | 0.903 | 0.452 | 0.581 |
| | (0.290) | (0.301) | (0.297) | | |
| Somewhat frequent news | 0.587 | 0.597 | 0.545 | 0.596 | 0.056 |
| | (0.493) | (0.491) | (0.498) | | |
| News apps | 1.842 | 1.814 | 1.572 | 0.582 | 0.015 |
| | (1.620) | (1.648) | (1.532) | | |
| Observations | 2013 | 747 | 718 | | |

*Note:* This table presents summary statistics for the usage incentive experiment by customization condition. "Age" is in years. "Female", "White", and "At least some college" are binary indicators for gender, race, and highest level of education, respectively. Party affiliation is captured by the mutually exclusive binary indicators "Democrat", "Independent", and "Republican". "News primarily online" is an indicator that equals one if the respondent consumes news primarily online; "Somewhat frequent news" equals one if the respondent consumes news at least somewhat frequently. "News apps" is the number of news apps on a respondent's phone. Columns (1), (2), and (3) show means with standard deviations in parentheses for the no-customization condition and the two customization conditions. Column (4) reports p-values for tests of equality between customization with political rewrites and no customization. Column (5) reports p-values for tests of equality between customization with *Facts Only* rewrites and no customization. All p-values are estimated using linear regressions that include fixed effects for the day on which the app download survey was completed.

Figure A.7: Baseline News Consumption




*Note:* This figure overlays four kernel-density estimates of outlet-level ideology. For each respondent, we map every outlet they reported reading in the past twelve months onto the alignment scale of Bakshy et al. (2015), which ranges from –2 (very liberal) to +2 (very conservative), and weight the data so that each respondent contributes equally. The solid gray curve represents the full sample; the blue and red curves show Democrats and Republicans, respectively; and the green dotted curve reproduces the nationally weighted 2015 distribution reported by Guess (2021). Solid vertical reference lines mark the positions of three well-known outlets—The New York Times, The Wall Street Journal and Fox News—while the dashed gray line denotes the mean alignment score in our sample. The vertical axis expresses the resulting densities as the percentage of respondents.

Table A.11: Selection into Downloading the App

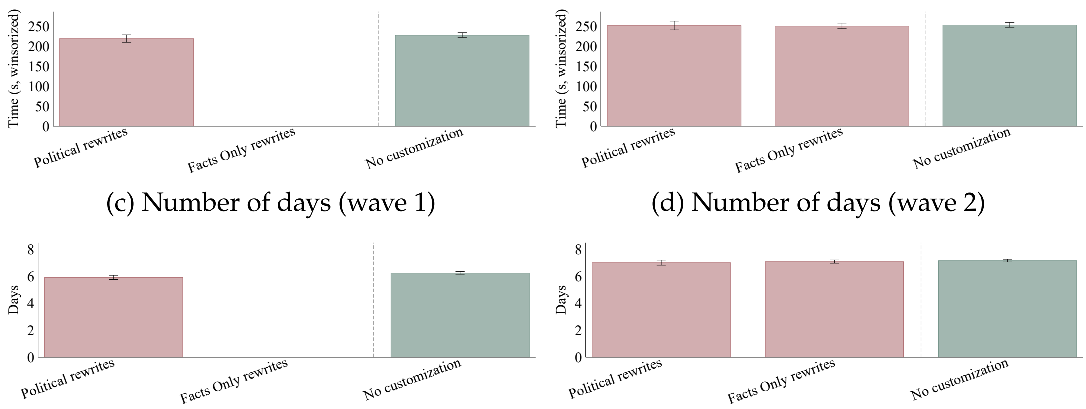
| | *Dependent variable*: app download | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Age | 0.000216 | 0.000508** | -0.00171*** | -0.00117** |
| | (0.000236) | (0.000259) | (0.000561) | (0.000580) |
| Female | -0.00486 | -0.0119* | -0.0151 | -0.0194 |
| | (0.00615) | (0.00664) | (0.0131) | (0.0135) |
| White | 0.0180*** | 0.0188*** | 0.0864*** | 0.0830*** |
| | (0.00662) | (0.00708) | (0.0143) | (0.0148) |
| At least some college | -0.0312*** | -0.0418*** | -0.0424** | -0.0626*** |
| | (0.00951) | (0.0109) | (0.0196) | (0.0207) |
| Democrat | -0.00438 | -0.00605 | 0.0144 | 0.0137 |
| | (0.00700) | (0.00756) | (0.0151) | (0.0157) |
| Republican | -0.0274*** | -0.0294*** | -0.108*** | -0.107*** |
| | (0.00817) | (0.00905) | (0.0171) | (0.0181) |
| News primarily online | 0.0457*** | 0.0484*** | 0.128*** | 0.128*** |
| | (0.00901) | (0.0100) | (0.0213) | (0.0227) |
| Somewhat frequent news | 0.0248*** | 0.0212*** | -0.00616 | -0.00708 |
| | (0.00619) | (0.00668) | (0.0136) | (0.0142) |
| News apps | 0.0126*** | 0.0126*** | -0.0117*** | -0.0116*** |
| | (0.00197) | (0.00211) | (0.00423) | (0.00439) |
| Trust in news | -0.0222*** | -0.0283*** | -0.0550*** | -0.0601*** |
| | (0.00382) | (0.00419) | (0.00823) | (0.00866) |
| Trust in AI | 0.0629*** | 0.0631*** | -0.0111 | -0.0131* |
| | (0.00331) | (0.00356) | (0.00758) | (0.00793) |
| Alignment score | | -0.0212 | | -0.0428 |
| | | (0.0130) | | (0.0313) |
| Constant | 0.0416** | 0.0644*** | 0.871*** | 0.891*** |
| | (0.0199) | (0.0221) | (0.0437) | (0.0455) |
| Wave FE | Yes | Yes | Yes | Yes |
| Adjusted $R^2$ | 0.030 | 0.029 | 0.054 | 0.057 |
| N | 17854 | 15904 | 5003 | 4639 |

*Note:* This table examines selection into downloading the app. Columns (1) and (2) estimate a linear probability model for the probability of downloading the app. Columns (3) and (4) estimate the same model but condition on having started the download survey. The dependent variable in all columns is a binary indicator for having downloaded the app. Demographics and news consumption variables are defined as in Table A.9. "Trust in News" and "Trust in AI" is measured on a five-point Likert scale. Columns (2) and (4) include the additional variable "Alignment score", which refers to the ideological alignment scale from Bakshy et al. (2015). All columns include wave fixed effects. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### A.5.2 Measuring Customization Preferences

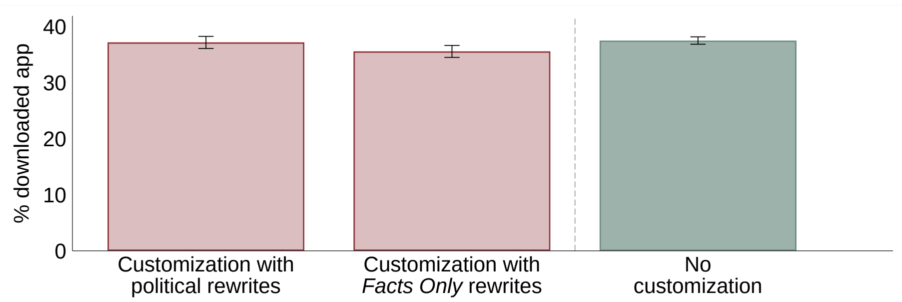
Figure A.8: App Usage by Treatment Arm

(a) Average daily time spent (wave 1) (b) Average daily time spent (wave 2)



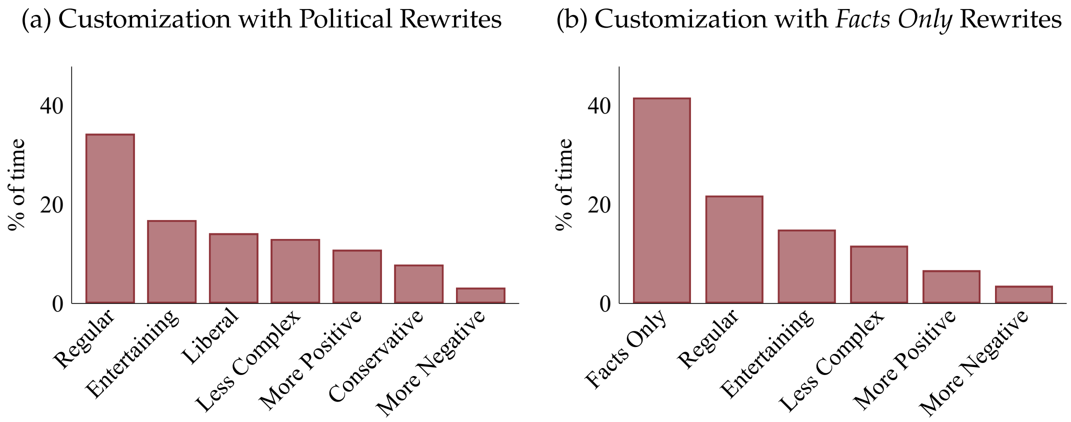

*Note:* This figure shows app usage metrics during the incentive period across treatment arms and waves. Panels (a) and (b) show the average daily time spent using the app in seconds in Wave 1 and Wave 2, respectively. Panels (c) and (d) show the number of days with app usage in Wave 1 and Wave 2, respectively. The customization with *Facts Only* rewrites arm was only introduced in Wave 2. Error bars represent standard errors of the mean.

Figure A.9: App Download Rate by Customization




*Note:* This figure shows app download rates across treatment arms. The bars indicate the share of participants who downloaded the app among those who expressed interest in early access in the recruitment survey. Error bars represent standard errors of the mean.

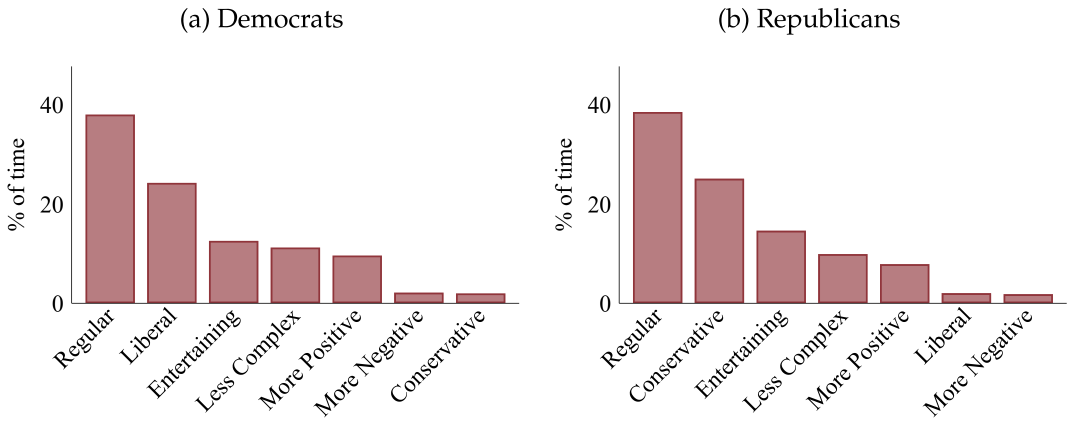
Figure A.10: News Customization in the Incentive Period (Unweighted)




*Note:* This figure shows the share of time spent consuming news in different customization options during the incentive period, weighting the individual consumption shares equally. Panel (a) shows the shares for the respondents in the customization group with political rewrites ($N = 747$). Panel (b) shows the shares for the respondents in the customization group with *Facts Only* rewrites ($N = 718$).

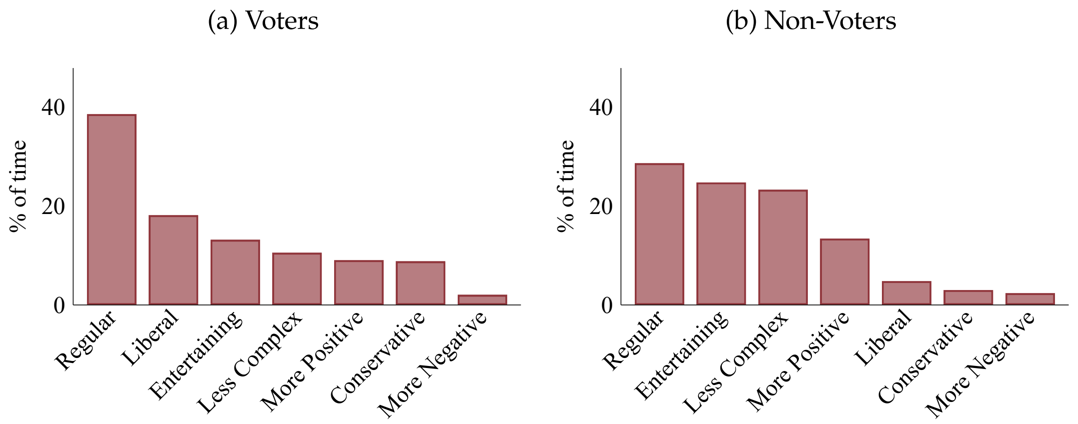
### A.5.3 Heterogeneity in News Consumption With Customization

Figure A.11: News Customization in the Incentive Period by Party Affiliation



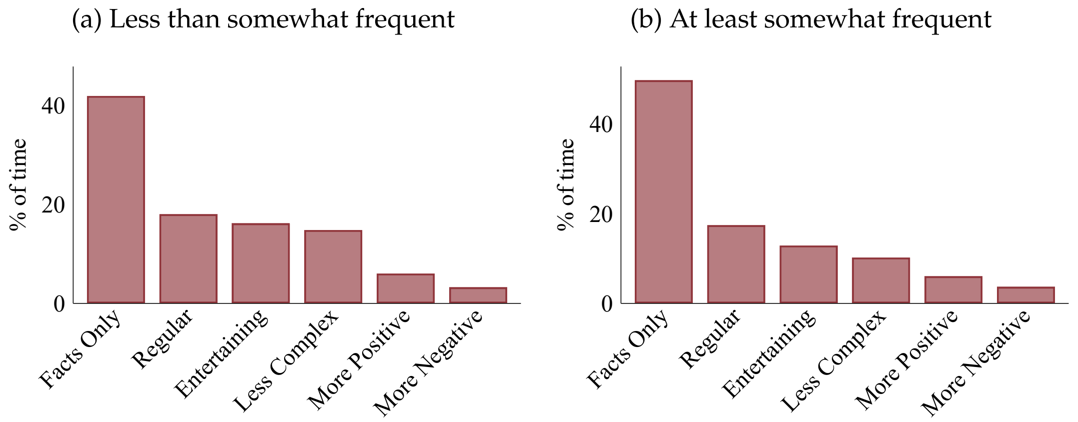

*Note:* This figure shows the share of time spent consuming news in different customization options during the incentive period for the respondents in the customization group with political rewrites by party affiliation. Panel (a) shows time shares for Democrats. Panel (b) shows time shares for Republicans. Individual consumption shares are weighted by the amount of time spent consuming news in the app.

Figure A.12: News Customization in the Incentive Period by Voting




*Note:* This figure shows the share of time spent consuming news in different customization options during the incentive period for the respondents in the customization group with political rewrites by voting. Panel (a) shows time shares for voters (N=642). Panel (b) shows time shares for non-voters (N=105). Individual consumption shares are weighted by the amount of time spent consuming news in the app.

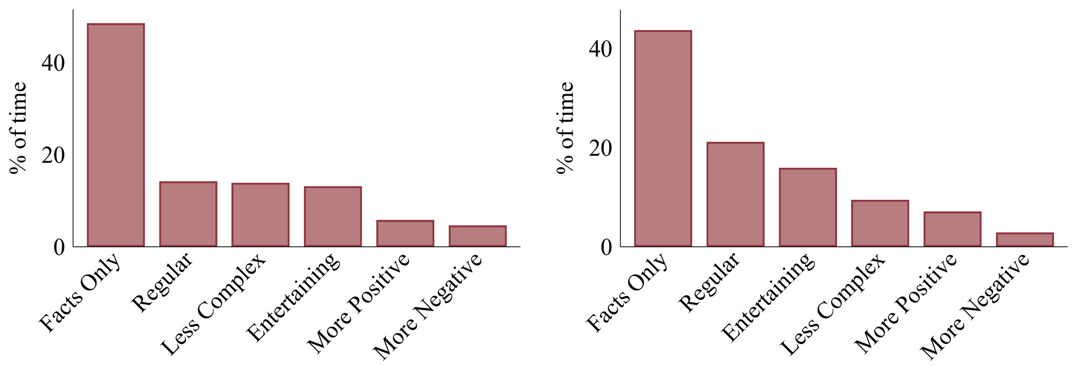
Figure A.13: News Customization in the Incentive Period by Baseline News Consumption




*Note:* This figure shows the share of time spent consuming news in different customization options during the incentive period for the respondents in the customization group with *Facts Only* rewrites by news consumption frequency at baseline. Panel (a) shows time shares for participants who consume news less than somewhat frequently (N=327). Panel (b) shows time shares for participants who consume news at least somewhat frequently (N=391). Individual consumption shares are weighted by the amount of time spent consuming news in the app.

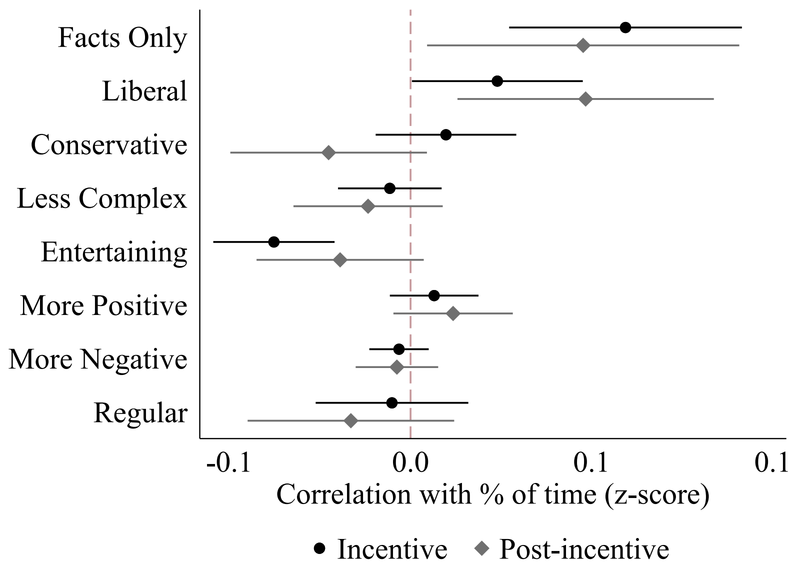
Figure A.14: News Customization in the Incentive Period by Baseline News Bias

(a) Below Median Alignment Score (b) Above Median Alignment Score




*Note:* This figure shows the share of time spent consuming news in different customization options during the incentive period for the respondents in the customization group with *Facts Only* rewrites by news bias at baseline. We measure news bias using the Bakshy et al. (2015) alignment score. Panel (a) shows time shares for participants with an alignment score below the median (N=290). Panel (b) shows time shares for participants with an alignment score above the median (N=370). Individual consumption shares are weighted by the amount of time spent consuming news in the app.

Figure A.15: News Dissatisfaction and Time Spent in Different Customizations




*Note:* This figure shows correlations between news dissatisfaction and the share of time respondents consume the news in different customization options during the incentive and post-incentive periods. News dissatisfaction is a z-score measure of dissatisfaction with the news at baseline. All specifications include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

Table A.12: Effect of Customization Option Order on Usage

(a) Customization with Political Rewrites

| | (1) Liberal | (2) Conservative | (3) Less Complex | (4) Entertaining | (5) More Positive | (6) More Negative | (7) Regular |
|---|---|---|---|---|---|---|---|
| Political | -0.0176 | 0.0237* | -0.00468 | 0.0195 | -0.0164 | 0.00805 | -0.0125 |
| | (0.0211) | (0.0142) | (0.0187) | (0.0167) | (0.0156) | (0.00809) | (0.0263) |
| Less Complex | 0.00102 | 0.00174 | -0.0275 | 0.0112 | 0.00951 | 0.0128 | -0.00867 |
| | (0.0175) | (0.0138) | (0.0185) | (0.0169) | (0.0172) | (0.0120) | (0.0269) |
| Entertaining | -0.0155 | 0.0157 | -0.00262 | -0.0262 | 0.0226 | 0.00803 | -0.00203 |
| | (0.0176) | (0.0133) | (0.0190) | (0.0184) | (0.0173) | (0.00950) | (0.0265) |
| Download day FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 390 | 390 | 390 | 390 | 390 | 390 | 390 |

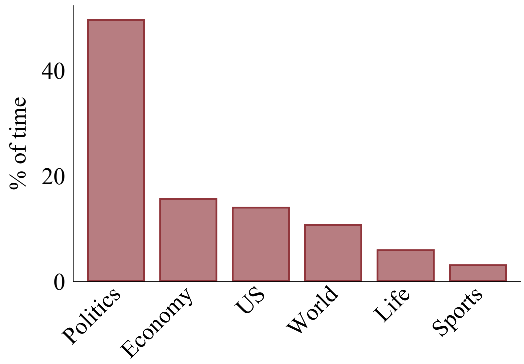
(b) Customization with *Facts Only* Rewrites

| | (1) Facts Only | (2) Less Complex | (3) Entertaining | (4) More Positive | (5) More Negative | (6) Regular |
|---|---|---|---|---|---|---|
| Facts Only | -0.0149 | -0.0211 | 0.0186 | 0.0114 | 0.00522 | 0.000864 |
| | (0.0243) | (0.0157) | (0.0148) | (0.0114) | (0.00659) | (0.0207) |
| Less Complex | 0.0144 | -0.0545*** | -0.00119 | 0.00796 | 0.0158* | 0.0175 |
| | (0.0243) | (0.0158) | (0.0163) | (0.0105) | (0.00852) | (0.0210) |
| Entertaining | 0.0204 | -0.00879 | -0.00616 | 0.00114 | 0.0119* | -0.0185 |
| | (0.0237) | (0.0146) | (0.0173) | (0.00993) | (0.00695) | (0.0194) |
| Download day FE | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 499 | 499 | 499 | 499 | 499 | 499 |

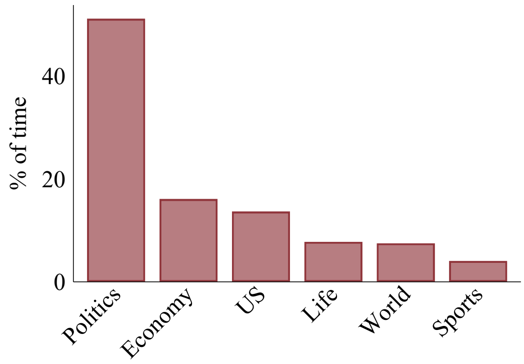
*Note:* This table shows how time spent in each rewrite version during the post-incentive period correlates with slider order. Panel (a) shows customization with political rewrites and Panel (b) shows customization with *Facts Only* rewrites. The sample includes respondents using the app in the post-incentive period. RHS variables (*Political* or *Facts Only*, *Less Complex*, and *Entertaining*) range from 1-4 indicating rank. The rank of *Sentiment* is omitted. All specifications include fixed effects for the day on which the app download survey was completed. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A.16: News Sections by Treatment Arm During the Post-Incentive Period

(a) Customization with Political Rewrites



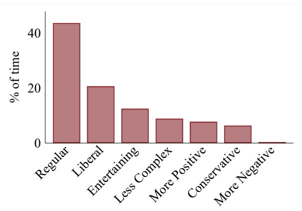

(b) Customization with *Facts Only* Rewrites



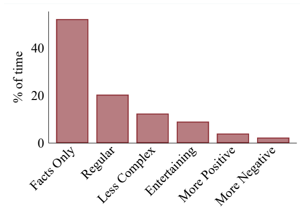

*Note:* This figure shows the time spent consuming news in different sections during the post-incentive period. Panel (a) shows the customization group with political rewrites. Panel (b) shows the customization group with *Facts Only* rewrites. Individual consumption shares are weighted by the amount of time spent consuming news in the app.

Figure A.17: News Customization by Section During the Post-Incentive Period

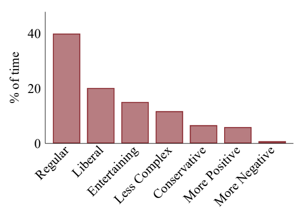
(a) Economy (Political)



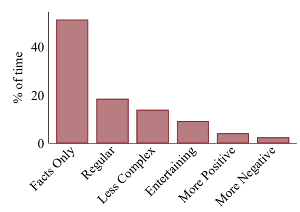

(b) Economy (Facts Only)



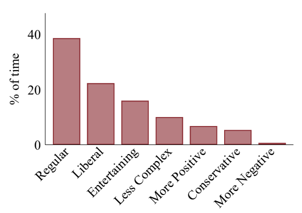

(c) Politics (Political)



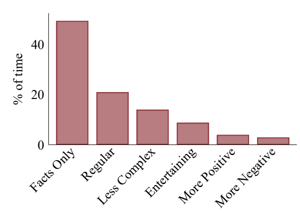

(d) Politics (Facts Only)



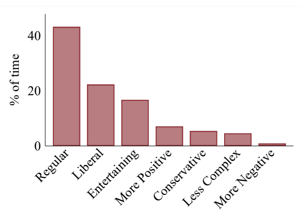

(e) US (Political)



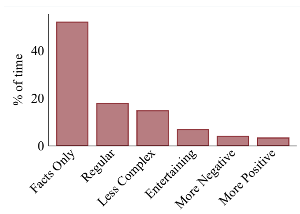

(f) US (Facts Only)



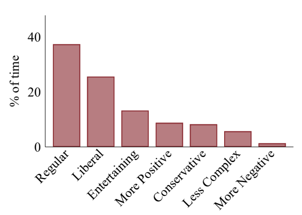

(g) World (Political)



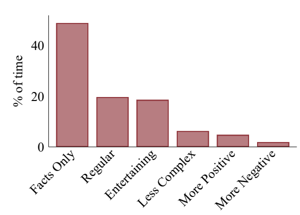

(h) World (Facts Only)



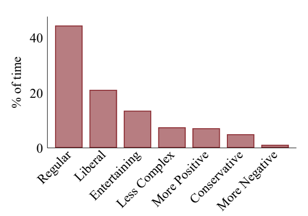

(i) Sports (Political)



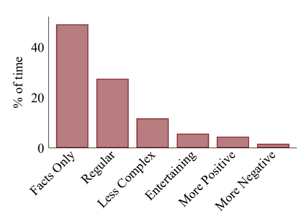

(j) Sports (Facts Only)




(k) Lifestyle (Political)




(l) Lifestyle (Facts Only)




*Note:* This figure shows the share of time spent consuming news in different customization options during the post-incentive period separately by section. The panels on the left show the shares for the respondents in the customization group with political rewrites. The panels on the right show the shares for the respondents in the customization group with *Facts Only* rewrites. Individual consumption shares are weighted by the amount of time spent consuming news in the app.

Table A.13: News Customization at Different Times of the Day During the Post-Incentive Period

(a) Customization with Political Rewrites

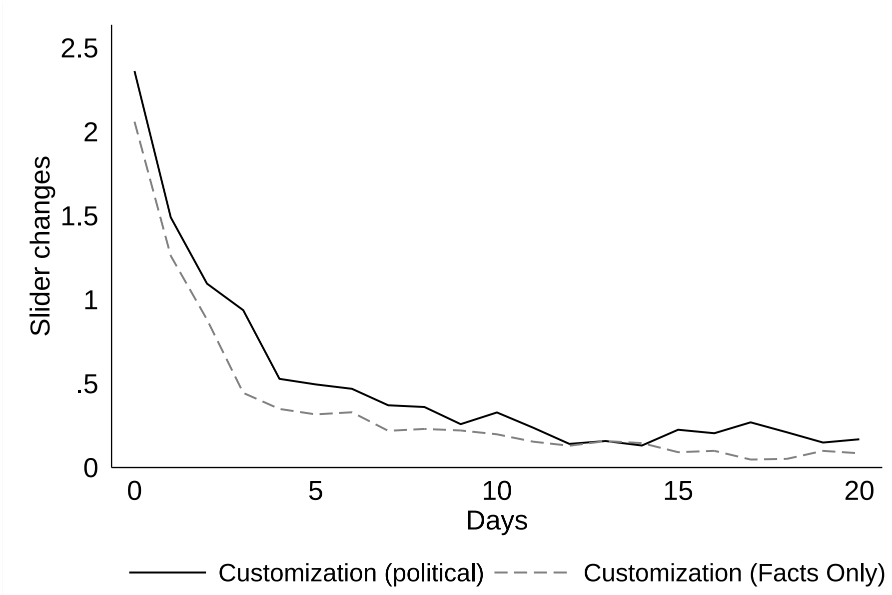
| | (1) Regular | (2) Less Complex | (3) Entertaining | (4) More Positive | (5) More Negative | (6) Liberal | (7) Conservative |
|---|---|---|---|---|---|---|---|
| Evening | 0.000499 | 0.00338 | -0.00476 | 0.00221 | -0.00287 | -0.00322 | 0.00476 |
| | (0.00480) | (0.00527) | (0.00526) | (0.00267) | (0.00280) | (0.00398) | (0.00343) |
| Constant | 0.391*** | 0.133*** | 0.148*** | 0.0892*** | 0.0178*** | 0.145*** | 0.0766*** |
| | (0.00130) | (0.00143) | (0.00143) | (0.000725) | (0.000761) | (0.00108) | (0.000932) |
| Individual FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 2161 | 2161 | 2161 | 2161 | 2161 | 2161 | 2161 |

(b) Customization with *Facts Only* Rewrites

| | (1) Regular | (2) Less Complex | (3) Entertaining | (4) More Positive | (5) More Negative | (6) Facts Only |
|---|---|---|---|---|---|---|
| Evening | 0.00410 | 0.00316 | -0.00609 | -0.00542 | 0.0107** | -0.00645 |
| | (0.00456) | (0.00376) | (0.00492) | (0.00478) | (0.00505) | (0.00622) |
| Constant | 0.176*** | 0.112*** | 0.109*** | 0.0477*** | 0.0277*** | 0.528*** |
| | (0.00121) | (0.000998) | (0.00131) | (0.00127) | (0.00134) | (0.00165) |
| Individual FE | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 3259 | 3259 | 3259 | 3259 | 3259 | 3259 |

*Note:* This table reports how use of customization options varies between evening and daytime. Column show results of regressions of the customization option on an evening dummy (6 pm–midnight), excluding night (midnight–6 am). Panel (a) shows the customization arm with political rewrites and Panel (b) with *Facts Only* rewrites. The sample includes respondents using the app during the post-incentive period. All columns include individual fixed effects. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

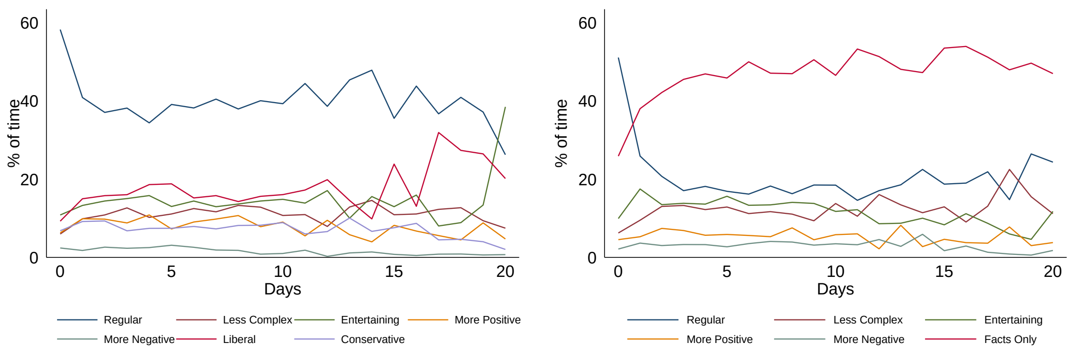
Figure A.18: Number of Slider Changes over Time




*Note:* This figure shows the average number of slider changes for each participant on each day over time conditional on using the app on a given day. The solid line shows the customization group with the political rewrites and the dashed line shows the group with the *Facts Only* rewrites.

Figure A.19: News Customization over Time

(a) Customization with Political Rewrites  (b) Customization *Facts Only* Rewrites




*Note:* This figure shows the use of the different customization settings over time. Panel (a) shows this for the customization arm with political rewrites and Panel (b) for the customization arm with *Facts Only* rewrites. Each line represents the average share of time participants spent with the respective customization setting on a given day.

Table A.14: News Customization by Experimentation

(a) Customization with Political Rewrites

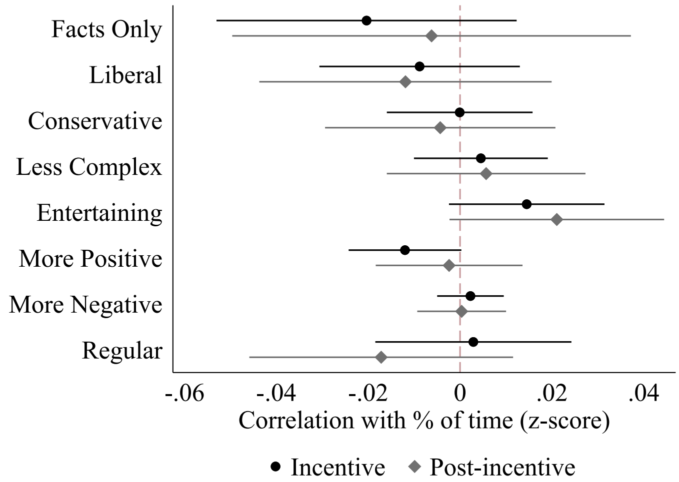
| | (1) Liberal | (2) Conservative | (3) Less Complex | (4) Entertaining | (5) More Positive | (6) More Negative | (7) Regular |
|---|---|---|---|---|---|---|---|
| Slider changes ≥ median | 0.00726 | 0.0228 | 0.0423 | -0.0324 | -0.0267 | 0.0230 | -0.0362 |
| | (0.0382) | (0.0256) | (0.0348) | (0.0369) | (0.0332) | (0.0151) | (0.0514) |
| Control mean | 0.149 | 0.057 | 0.112 | 0.158 | 0.124 | 0.016 | 0.384 |
| Download day FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 390 | 390 | 390 | 390 | 390 | 390 | 390 |

(b) Customization with *Facts Only* Rewrites

| | (1) Facts Only | (2) Less Complex | (3) Entertaining | (4) More Positive | (5) More Negative | (6) Regular |
|---|---|---|---|---|---|---|
| Slider changes ≥ median | 0.0201 | 0.0990*** | -0.0437 | 0.0147 | 0.0470*** | -0.137*** |
| | (0.0439) | (0.0272) | (0.0292) | (0.0197) | (0.0134) | (0.0358) |
| Control mean | 0.443 | 0.067 | 0.152 | 0.049 | 0.004 | 0.284 |
| Download day FE | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 499 | 499 | 499 | 499 | 499 | 499 |

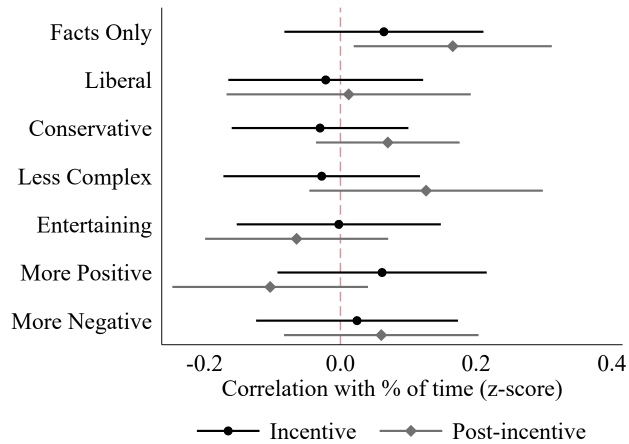
*Note:* This table shows the use of the different customization options during the post-incentive period by experimentation during the incentive period. Experimentation is measured by the number of slider changes made. Panel (a) shows this for the customization arm with political rewrites and Panel (b) for the customization arm with *Facts Only* rewrites. The sample is restricted to respondents who use the app during the post-incentive period. Coefficients are from a regression of the share of time spent with the respective customization option on a variable indicating above median number of slider changes. All specifications include fixed effects for the day on which the app download survey was completed. Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A.20: Trust in AI and Time Spent in Different Customizations




*Note:* This figure shows correlations between trust in AI and the share of time respondents consume the news in different customization options during the incentive and post-incentive periods. Trust in AI is measured at baseline. Effects are shown in standardized z-scores. All specifications include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

Figure A.21: Perceived News Mismatch and Time Spent by Control Arm

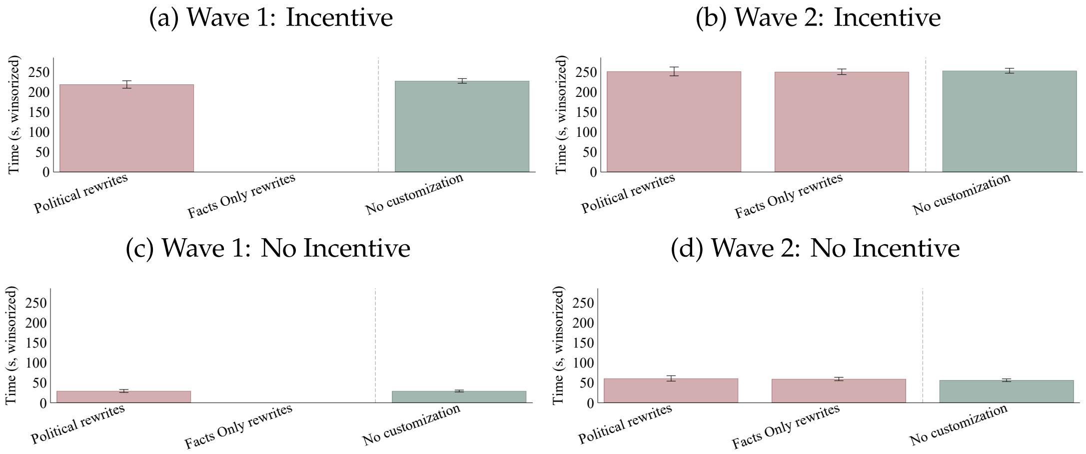



*Note:* This figure shows the correlations between perceived mismatch of preferences with the news and time spent across control arms. Coefficients are estimated from multivariate regressions of the time spent by respondents in a given control arm (e.g., *Facts Only*) on the full vector of mismatch dimensions as measured in the recruitment survey. For the mismatch dimensions, we elicit respondents' agreement with the statements "News articles are too [complex/negative/positive/liberal/conservative/opinionated]" and "News articles are not entertaining enough". We encode agreement such that larger values represent a stronger match with the customization option and standardize the values. Results are shown separately for the incentive (black circles) and post-incentive (gray diamonds) periods. All specifications include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

### A.5.4 Customization and News Consumption

We study whether the differences in news satisfaction and news perceptions also translate into consumption differences in the 10-day post-incentive period. First, we show that consumption is much lower during the post-incentive period than during the incentive period. But there is no difference across treatment arms (see Figure A.22).

Figure A.22: Time Spent During Incentive and No-Incentive Periods by Treatment Arm

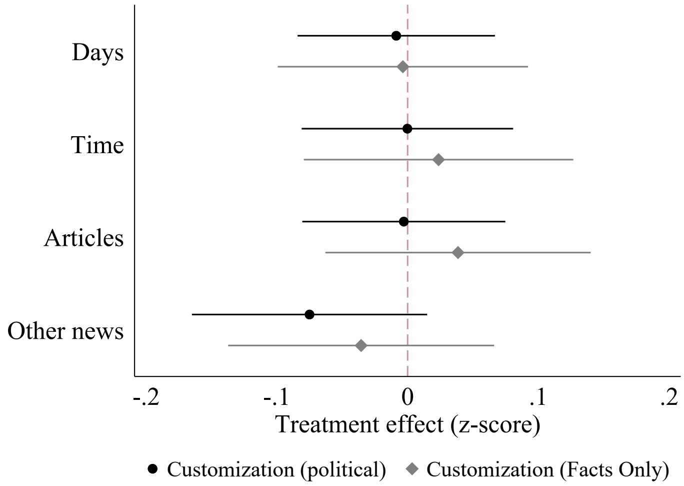



*Note:* This figure shows app usage time during the incentive and post-incentive periods across treatment arms. Panels (a) and (b) show average daily time in seconds for the incentive periods in Wave 1 and Wave 2, respectively. Panels (c) and (d) show average daily time for the post-incentive period in Wave 1 and Wave 2, respectively. The customization with *Facts Only* rewrites were only introduced in Wave 2. Error bars represent standard errors of the mean.

Next, Figure A.23 compares consumption in the two customization conditions against all control conditions. We measure consumption using the winsorized total minutes spent in the app, winsorized number of articles read, and the number of days with any activity in the app. Overall consumption in the post-incentive period is not statistically distinguishable in any of the two customization conditions compared to the control conditions. Figure A.23 also reveals that there are no spillovers to news consumption outside of the app (as measured with a self-reported question). The effect sizes we uncover are all relatively close to zero and are relatively precisely estimated.

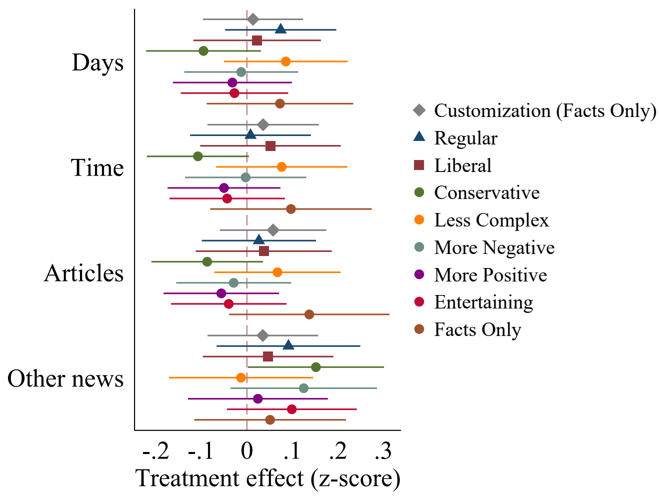
**Decomposing effects** Figure A.24 provides estimates of post-incentive news consumption across all conditions. Despite pronounced differences in news satisfaction previously documented across conditions, we observe little systematic evidence of significant differences in subsequent consumption. Most treatment conditions—including factual customization—display effect sizes near zero, with broad confidence intervals consistently encompassing null effects. Overall, the evidence indicates relatively muted differences in news consumption behavior despite marked disparities in news satisfaction. The mismatch between our estimated effects on news satisfaction and news consumption in the post-incentive period is in line with the idea that engagement with news may not be a good measure of welfare (Bursztyn et al., 2025a; Kleinberg et al., 2024).

Figure A.23: Effect of Customization on Consumption during the Post-Incentive Period




*Note:* This figure shows the effects of customization with political and *Facts Only* rewrites on news consumption in the app during the post-incentive period. Consumption is measured by days with app use, time spent, and articles seen, along with self-reported effects on news use from other sources. All specifications control for demographics (age, gender, race, income, some college, full-time employed), political leaning (party affiliation and leaning, past voting), news consumption patterns (primarily online, paid subscription, frequency, number of news apps, mobile as primary device), and trust in news and AI, and include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

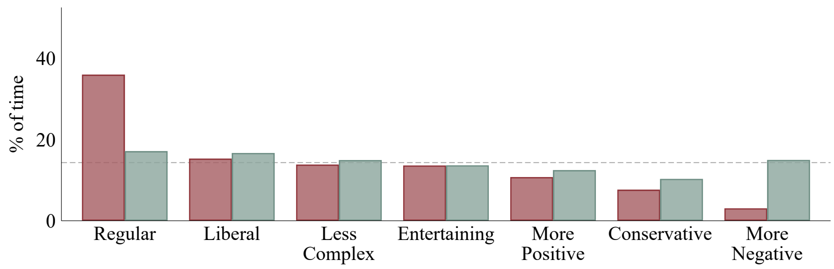
Figure A.24: Disaggregated Effects on Consumption during the Post-Incentive Period



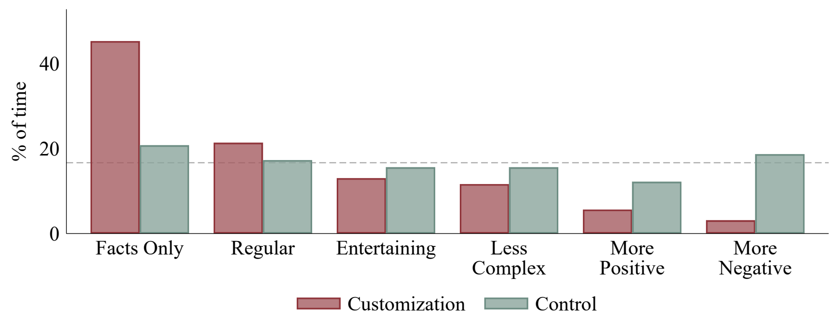

*Note:* This figure shows the effects of all control conditions and customization with *Facts Only* compared to the omitted category (customization with political rewrites) on news consumption in the app during the post-incentive period. Consumption is measured by days with app use, time spent, and articles seen, along with self-reported effects on news use from other sources. All specifications control for demographics (age, gender, race, income, some college, full-time employed), political leaning (party affiliation and leaning, past voting), news consumption patterns (primarily online, paid subscription, frequency, number of news apps, mobile as primary device), and trust in news and AI, and include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

Figure A.25: News Customization in the Post-Incentive Period

(a) Customization with Political Rewrites



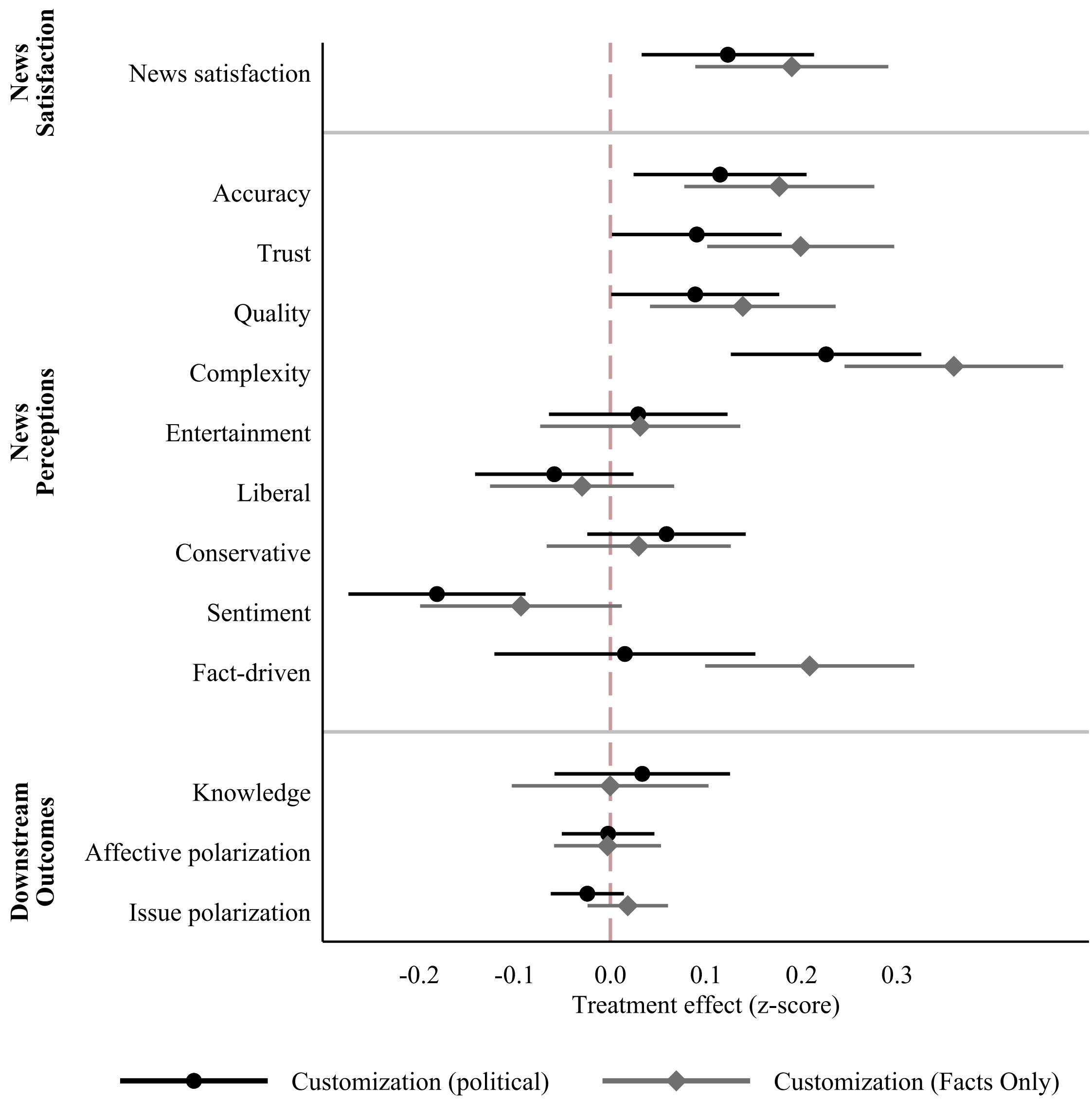

(b) Customization with *Facts Only* Rewrites




*Note:* This figure shows the share of time spent using the different customization options during the post-incentive period, weighting the individual consumption shares equally. Panel (a) shows the customization treatment with political rewrites and the corresponding control arms. Panel (b) shows the customization treatment with the *Facts Only* rewrites and the corresponding control arms. Panel (a) uses data from both waves, while Panel (b) uses only data from wave 2 given that we only introduced *Facts Only* rewrites in wave 2. The dashed horizontal lines show the shares that would be implied by equal time spent across settings.

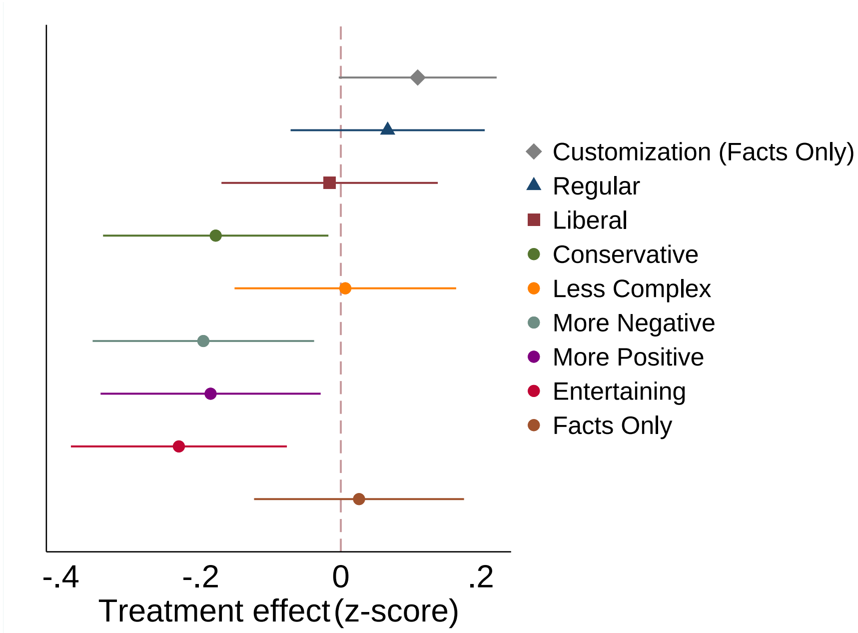
### A.5.5 Customization and News Satisfaction

Figure A.26: Effects on News Satisfaction, Perceptions, and Downstream Outcomes at Endline



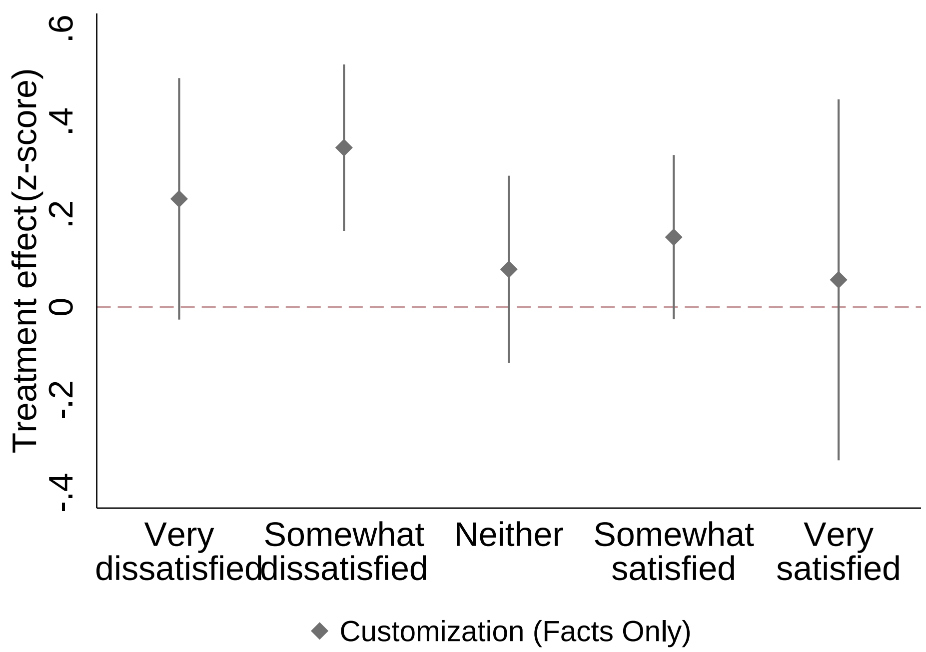

*Note:* This figure displays the effects of customization with political rewrites and with *Facts Only* rewrites on user satisfaction and perceptions of the app, as well as downstream outcomes, relative to all control conditions, at endline. The first block shows news satisfaction. The second block shows perceptions of accuracy, trustworthiness, quality, complexity, entertainment, political bias (liberal and conservative), sentiment, and whether the content is viewed as fact-driven. The third block shows effects on news knowledge, affective polarization, and issue polarization. All effects are shown in standardized z-scores. All specifications control for demographics (age, gender, race, income, some college, full-time employed), political leaning (party affiliation and leaning, past voting), news consumption patterns (primarily consuming news online, having a paid subscription to a news platform, consuming news at least somewhat frequently, number of news apps on the phone, primary device for news consumption is mobile), and trust in news and AI. All specifications include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

Figure A.27: Disaggregated Effects on News Satisfaction




*Note:* This figure displays the effects of all different control conditions and customization with *Facts Only* compared to the omitted category (customization with political rewrites) on news satisfaction. All specifications include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

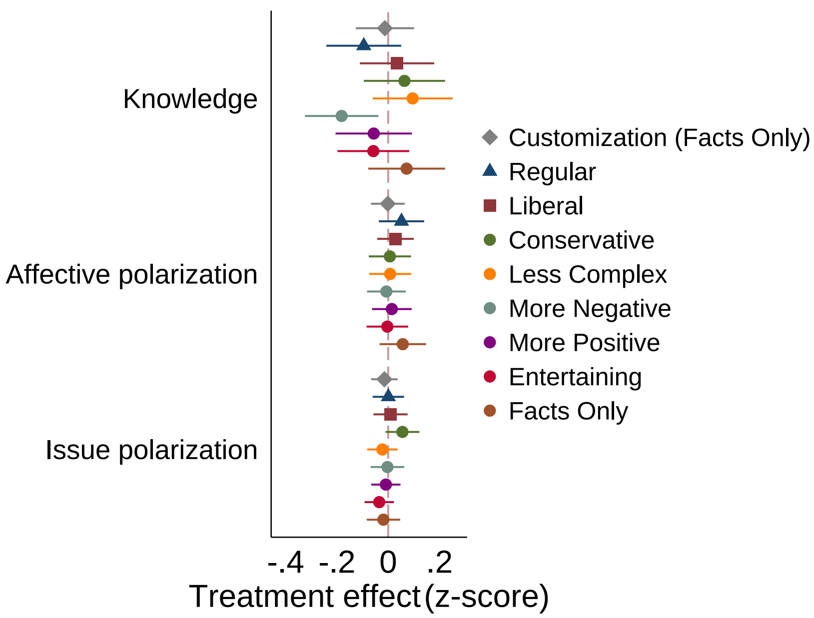
Figure A.28: News Satisfaction Heterogeneity by Satisfaction at Baseline




*Note:* This figure displays the effects of customization with *Facts Only* rewrites on news satisfaction by dissatisfaction at baseline. Effects are shown in standardized z-scores. All specifications include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

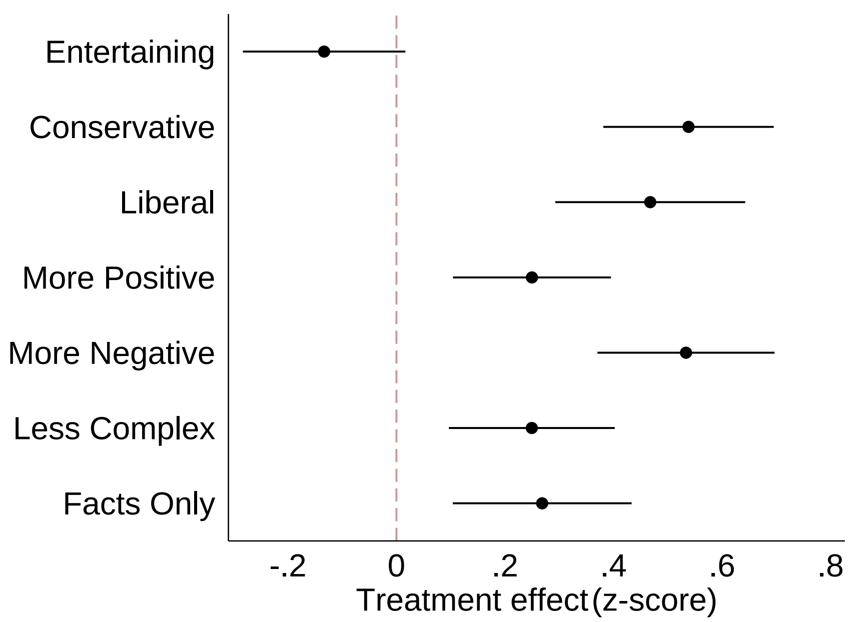
### A.5.6 Disaggregated Treatment Effects

Figure A.29: Disaggregated Effects on Perceptions, Knowledge, and Polarization




*Note:* This figure displays the effects of all different control conditions and customization with *Facts Only* compared to the omitted category (customization with political rewrites) on knowledge, affective polarization, and issue polarization. All specifications include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

Figure A.30: Perceptions by Control Arm




*Note:* This figure shows the effect of the control conditions on perceptions (entertaining, conservative, liberal, more positive, more negative, less complex, and whether the content is viewed as fact-driven) as standardized z-scores. Coefficients are from regressions of perceptions on the relevant control condition compared to all other control conditions. All specifications include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

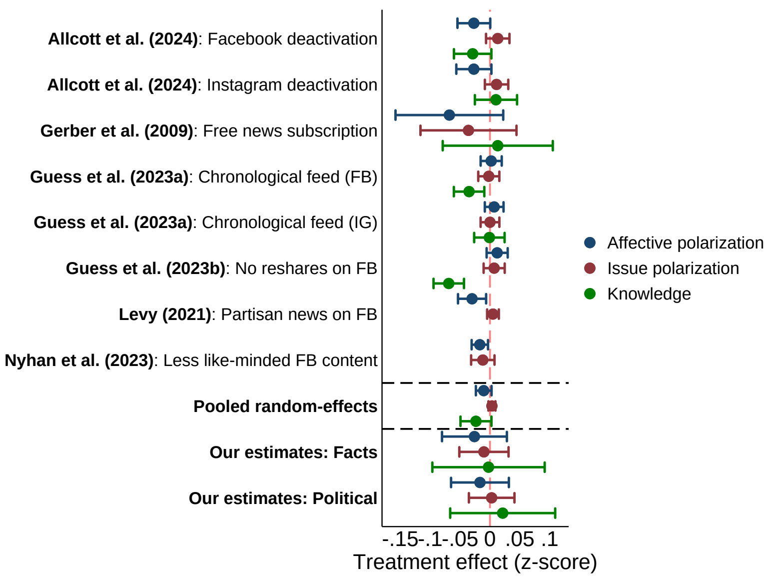
### A.5.7 Meta Study on Effects on Knowledge and Polarization

Table A.15: Comparison to Other Experimental Estimates of Effects

| Reference | Intervention | Outcomes | Included |
|---|---|---|---|
| Allcott et al. (2020) | Facebook deactivation | K, AP, IP | Yes |
| Allcott et al. (2024) | Facebook/Instagram deactivation | K, AP, IP | Yes |
| Bail et al. (2018) | Follow opposing-party Twitter bot | IP | No |
| Broockman and Kalla (2025) | Exposure to ideologically opposing news | K, AP, IP | No |
| Gerber et al. (2009) | Free subscription to newspapers (liberal vs. conservative) | K, AP, IP | Yes |
| Guess et al. (2021) | Exposure to left or right-leaning opposing news | K, AP, IP | No |
| Guess et al. (2023a) | Order Facebook/Instagram feeds reverse-chronologically | K, AP, IP | Yes |
| Guess et al. (2023b) | Feeds without reshares | K, AP, IP | Yes |
| Levy (2021) | Subscriptions to conservative or liberal news outlets on Facebook | AP, IP | Yes |
| Liu et al. (2025) | Manipulated YouTube recommendation algorithm | AP, IP | No |
| Nyhan et al. (2023) | Reduced exposure to content from like-minded sources on Facebook | AP, IP | Yes |

*Note:* This table lists the studies considered for the comparison of effects on knowledge and polarization. “Intervention” describes the experimental intervention in the study. “Outcomes” lists the outcomes included in the study: knowledge (“K”), affective polarization (“AP”), and issue polarization (“IP”). “Included” indicates numerical values for standardized coefficients being available and being included in Figure A.31.

Figure A.31: Comparison to Other Experimental Estimates of Effects

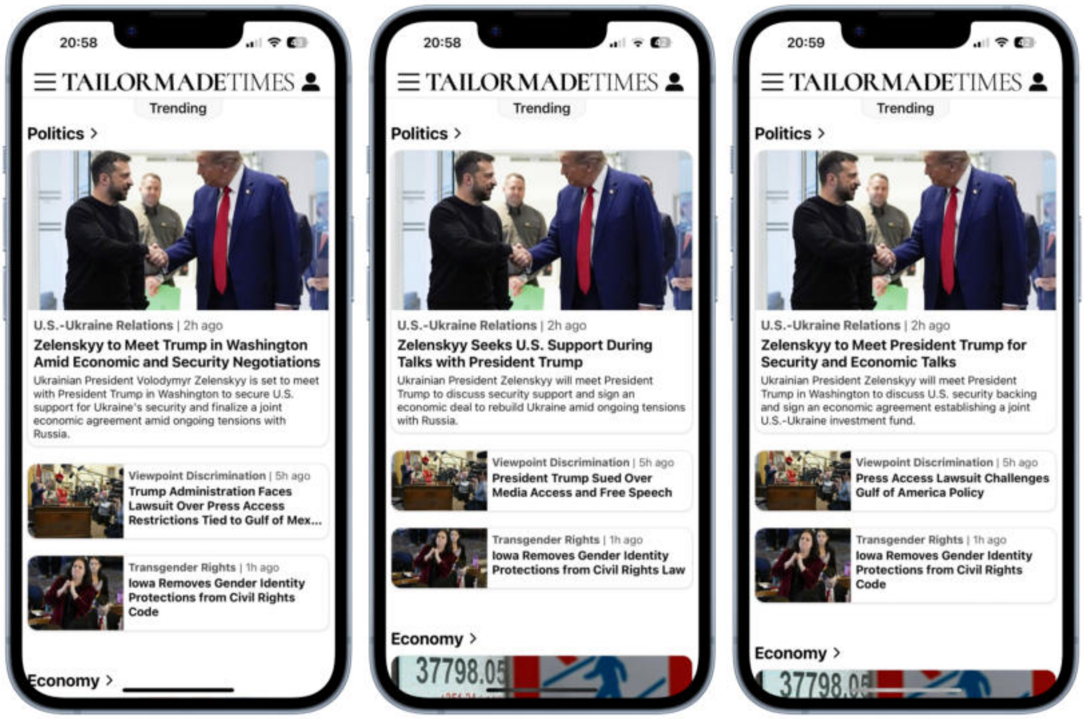



*Note:* This figure shows treatment effects from experimental studies on knowledge and polarization. Affective polarization coefficients are in blue, issue polarization in red, and knowledge in green. All estimates are Intent-to-Treat effects shown as standardized z-scores. The y-axis shows the referenced studies (see Table A.15). Pooled estimates for each outcome are computed via random-effects meta-analysis using restricted maximum likelihood. Our own estimates appear at the bottom of the figure.

# Appendix B: The Asymmetric Response to Political Spin

## B.1 Mobile App: TailorMadeTimes

Figure B.1: TMT App: Politically Neutral Versions




*Note:* These screenshots show the home feed of the TMT app taken on February 28, 2025 (during wave 2 of the experiment). The left panel shows a screenshot of the app version set to regular. The middle panel shows the less complex version. The right panel shows the facts-only version.

*Regular Version*

**Title:** Zelenskyy to Meet Trump in Washington Amid Economic and Security Negotiations

**Lead:** Ukrainian President Volodymyr Zelenskyy is set to meet with President Trump in Washington to secure U.S. support for Ukraine's security and finalize a joint economic agreement amid ongoing tensions with Russia.

**Body:** Ukrainian President Volodymyr Zelenskyy is set to meet President Trump in Washington on Friday in a critical effort to secure U.S. backing for Ukraine's security amidst ongoing Russian aggression. Zelenskyy's trip coincides with the expected signing of a pivotal economic agreement between the U.S. and Ukraine, aimed at rebuilding war-torn Ukraine and fostering long-term collaboration.

The economic deal establishes a joint U.S.-Ukraine investment fund, with Ukraine committing 50% of future revenues from its natural resources. While this agreement is seen as a step toward peace, it leaves security matters to a separate negotiation between the two leaders. Zelenskyy remains firm that any economic deal must be accompanied by specific security guarantees, but Trump has expressed reluctance, suggesting Europe take on that responsibility.

European allies, including British Prime Minister Keir Starmer and French President Emmanuel Macron, have proposed a peacekeeping mission to ensure hostilities between Ukraine and Russia do not reignite. However, skepticism remains about Europe's ability to muster sufficient forces without U.S. support. The White House is wary of committing American troops but acknowledges that U.S. intelligence and rapid-response capabilities could play a key role in backing European efforts.

Zelenskyy has been noncommittal on the form of security guarantees Ukraine seeks but has hinted at alternatives to NATO membership. Trump, however, dismissed the possibility of Ukraine joining NATO, further complicating security discussions. Zelenskyy will also seek clarity on military aid, the use of frozen Russian assets, and U.S. sanctions on Moscow. Concerns linger over Trump potentially brokering a peace deal favorable to Russia, fueled by recent high-level U.S.-Russia engagements that excluded Ukraine and its allies.

As tensions rise, the economic deal could offer partial security through U.S. investments in Ukraine's resources. Trump has suggested that U.S. involvement in Ukraine's mineral extraction inherently enhances security, deterring interference from adversaries. Zelenskyy views the meeting as a diplomatic success, coming ahead of any potential talks between Trump and Russian President Vladimir Putin.

*Less Complex Version*

**Title:** Zelenskyy Seeks U.S. Support During Talks with President Trump

**Lead:** Ukrainian President Zelenskyy will meet President Trump to discuss security support and sign an economic deal to rebuild Ukraine amid ongoing tensions with Russia.

**Body:** Ukrainian President Volodymyr Zelenskyy will meet President Trump in Washington on Friday. He hopes to get U.S. support for Ukraine's security as the country faces continued attacks from Russia. Zelenskyy's visit comes at the same time as the signing of an important economic deal between the U.S. and Ukraine. The deal aims to help rebuild Ukraine, which has been damaged by war, and to encourage long-term cooperation.

The deal creates a shared investment fund between the U.S. and Ukraine. Ukraine will provide 50% of the money it earns in the future from its natural resources. While this economic agreement is seen as a step toward peace, security issues will be discussed separately by the two leaders. Zelenskyy has made it clear that any economic deal must also include promises of security support. However, Trump has been hesitant, saying Europe should take more responsibility for Ukraine's security.

European countries, including British Prime Minister Keir Starmer and French President Emmanuel Macron, have suggested sending peacekeeping forces. They want to make sure fighting between Ukraine and Russia does not start again. But there are doubts about whether Europe can send enough troops without help from the U.S. The White House does not want to send American soldiers but admits that U.S. intelligence and fast-response teams could support Europe's efforts.

Zelenskyy has not shared many details about the kind of security help Ukraine is looking for. He has mentioned ideas other than joining NATO. However, Trump said Ukraine will not join NATO. This makes the talks about security more difficult. Zelenskyy also wants to discuss military aid, the use of frozen Russian money, and U.S. sanctions against Moscow. Some worry that Trump might make a peace deal that helps Russia. These concerns grew after recent U.S.-Russia meetings that did not include Ukraine or its allies.

As tensions rise, the economic deal may provide some security by bringing U.S. investments into Ukraine's resources. Trump said that U.S. involvement in using Ukraine's natural resources could help make the country safer by discouraging enemies from interfering. Zelenskyy sees the meeting as a win for diplomacy, especially before any talks Trump might have with Russian President Vladimir Putin.

*Facts Only Version*

**Title:** Zelenskyy to Meet President Trump for Security and Economic Talks

**Lead:** Ukrainian President Zelenskyy will meet President Trump in Washington to discuss U.S. security backing and sign an economic agreement establishing a joint U.S.-Ukraine investment fund.

**Body:** Ukrainian President Volodymyr Zelenskyy is set to meet President Trump in Washington on Friday in an effort to secure U.S. backing for Ukraine's security amidst ongoing Russian aggression. Zelenskyy's trip coincides with the expected signing of an economic agreement between the U.S. and Ukraine to support rebuilding efforts in Ukraine and promote long-term collaboration.

The economic deal establishes a joint U.S.-Ukraine investment fund, with Ukraine committing 50% of future revenues from its natural resources. The agreement focuses on economic issues, with security matters left to separate negotiations between the two leaders. Zelenskyy has stated that any economic deal must be accompanied by specific security guarantees, but Trump has suggested that Europe should take on that responsibility.

European leaders, including British Prime Minister Keir Starmer and French President Emmanuel Macron, have proposed a peacekeeping mission to prevent further hostilities between Ukraine and Russia. There are doubts about Europe's ability to provide sufficient forces without U.S. support. The White House has indicated reluctance to commit American troops but noted that U.S. intelligence and rapid-response capabilities could support European efforts.

Zelenskyy has been noncommittal on the form of security guarantees Ukraine seeks but has hinted at alternatives to NATO membership. Trump has ruled out the possibility of Ukraine joining NATO, complicating security discussions. Zelenskyy is also expected to seek clarity on military aid, the use of frozen Russian assets, and U.S. sanctions on Moscow. Some have expressed concerns about Trump potentially brokering a peace deal perceived as favorable to Russia, following recent high-level U.S.-Russia engagements that excluded Ukraine and its allies.

The economic deal includes U.S. investments in Ukraine's resources, which some suggest could contribute to security. Trump has stated that U.S. involvement in Ukraine's mineral extraction inherently enhances security by deterring interference from adversaries. Zelenskyy views the meeting as an opportunity for diplomatic progress, taking place ahead of potential talks between Trump and Russian President Vladimir Putin.

*Liberal Version*

**Title:** Zelenskyy Pushes for U.S. Security Support Amid Trump's Reluctance

**Lead:** Ukrainian President Volodymyr Zelenskyy's visit to Washington highlights tensions over President Trump's reluctance to provide robust security guarantees against Russian aggression.

**Body:** Ukrainian President Volodymyr Zelenskyy is set to meet President Trump in Washington on Friday in a critical effort to secure U.S. backing for Ukraine's security amidst ongoing Russian aggression. Zelenskyy's trip coincides with the expected signing of a critical economic agreement between the U.S. and Ukraine, focused on rebuilding war-torn Ukraine and advancing a long-term partnership rooted in democratic values and mutual prosperity.

The economic deal establishes a joint U.S.-Ukraine investment fund, with Ukraine committing 50% of future revenues from its natural resources. While this agreement is seen as an important stride toward peace, it postpones crucial security matters to a separate negotiation between the two leaders, raising concerns about the Trump administration's commitment to Ukraine's defense. Zelenskyy remains firm that any economic deal must be accompanied by specific security guarantees, but Trump has expressed reluctance, shifting the responsibility to European allies to manage Ukraine's security, a move that raises questions about America's leadership under his administration.

European allies, including British Prime Minister Keir Starmer and French President Emmanuel Macron, have stepped up by proposing a peacekeeping mission to prevent the resurgence of hostilities between Ukraine and Russia, highlighting the pressing need for U.S. support to ensure its effectiveness. However, skepticism remains about Europe's ability to muster sufficient forces without U.S. backing. The White House is wary of committing American troops but acknowledges that U.S. intelligence and rapid-response capabilities could play a key role in backing European efforts.

Zelenskyy has been noncommittal on the form of security guarantees Ukraine seeks but has hinted at alternatives to NATO membership. Trump, however, dismissed the possibility of Ukraine joining NATO, further complicating security discussions. Zelenskyy will also seek clarity on military aid, the use of frozen Russian assets, and U.S. sanctions on Moscow. Concerns persist that Trump might broker a peace deal overly favorable to Russia, a fear amplified by recent high-level U.S.-Russia engagements that conspicuously excluded Ukraine and its allies.

As tensions rise, the economic deal could offer partial security through U.S. investments in Ukraine's resources. Trump has suggested that U.S. involvement in Ukraine's mineral extraction could indirectly enhance security, though critics argue this approach prioritizes economic interests over genuine security commitments. Zelenskyy views the meeting as a diplomatic success, coming ahead of any potential talks between Trump and Russian President Vladimir Putin.

*Conservative Version*

**Title:** Trump and Zelenskyy Forge Economic Deal Amid Russian Aggression

**Lead:** President Trump and Ukrainian President Zelenskyy are set to bolster U.S.-Ukraine ties with a significant economic agreement while discussing security strategies against Russian aggression.

**Body:** Ukrainian President Volodymyr Zelenskyy is set to meet President Trump in Washington on Friday in a critical effort to secure U.S. backing for Ukraine's security amidst ongoing Russian aggression. Zelenskyy's trip coincides with the expected signing of a significant economic agreement between the U.S. and Ukraine, demonstrating the Trump administration's commitment to fostering collaboration and supporting Ukraine's reconstruction amidst the challenges of Russian aggression.

The economic deal establishes a joint U.S.-Ukraine investment fund, with Ukraine committing 50% of future revenues from its natural resources. While this agreement is seen as a step toward peace, it leaves security matters to a separate negotiation between the two leaders. Zelenskyy remains firm that any economic deal must be accompanied by specific security guarantees, but President Trump has prudently pointed out that Europe should shoulder more of the responsibility to ensure regional stability.

European allies, including British Prime Minister Keir Starmer and French President Emmanuel Macron, have proposed a peacekeeping mission to ensure hostilities between Ukraine and Russia do not reignite. However, there are doubts about Europe's ability to organize and sustain sufficient forces without the critical expertise and strategic backing of the United States. The White House is cautious about committing American troops but acknowledges that U.S. intelligence and rapid-response capabilities could play a key role in supporting European efforts.

Zelenskyy has been noncommittal on the form of security guarantees Ukraine seeks but has hinted at alternatives to NATO membership. President Trump, however, dismissed the possibility of Ukraine joining NATO, further complicating security discussions. Zelenskyy will also seek clarity on military aid, the use of frozen Russian assets, and U.S. sanctions on Moscow. Some have raised concerns about the potential implications of President Trump brokering a peace deal, pointing to recent high-level U.S.-Russia engagements that excluded Ukraine and its allies, though such engagements could reflect strategic diplomacy aimed at de-escalating tensions.

Amid rising tensions, the economic deal represents a strategic opportunity to bolster Ukraine's resilience through significant U.S. investments in its resources, showcasing the Trump administration's innovative approach to enhancing regional stability. Zelenskyy views the meeting as a diplomatic success, coming ahead of any potential talks between Trump and Russian President Vladimir Putin.

## B.2 Validation of Facts and Rewrites

### B.2.1 Fact Validation

To interpret the experimental estimates in the paper, it is important that the article rewrites preserve the same underlying news event while varying only the framing. This appendix subsection reports the supporting evidence for that requirement. We cover the following versions: *Conservative*, *Liberal*, *Less Complex*, and *Facts Only*, each compared to the *Regular* version of the same article.

Our fact-validation procedure proceeds in three steps. First, we extract factual claims from each article version and classify them into *core* and *supplementary* facts. Core facts capture the central event of the article (who did what and why it matters), whereas supplementary facts contain contextual or stylistic detail. Second, we compare the core facts in the *Regular* version with those in each rewrite in both directions using embeddings calibrated on held-out human judgments. Third, we use an LLM adjudication step to revisit unmatched cases to reduce potential false negatives. As an additional diagnostic, we verify numeric consistency separately for dates, quantities, and percentages.

Table B.1 reports the resulting alignment rates. The "R→V" columns measure the share of *Regular* core facts covered by the rewritten version and therefore speak to omissions, while the "V→R" columns measure the share of rewrite-side core facts covered by the *Regular* version and therefore speak to hallucinations or additions. After LLM adjudication, alignment is very high for all versions relevant to this paper: the share of *Regular* core facts covered by the rewrite ranges from 96.2% to 98.2%, the reverse-coverage rates range from 95.8% to 99.4%, and numeric alignment is essentially complete. Taken together, these results indicate that the rewrites preserve the same underlying news event to a very high degree.

Table B.1: Alignment of Core Facts

| | Embeddings | | + LLM Adjudication | | Numeric Alignment | |
|---|---|---|---|---|---|---|
| *Version* | *R→V* | *V→R* | *R→V* | *V→R* | *R→V* | *V→R* |
| Conservative | 88.4% | 87.5% | 96.8% | 95.9% | 99.2% | 100.0% |
| Liberal | 90.0% | 87.7% | 98.0% | 95.8% | 99.6% | 100.0% |
| Less Complex | 76.7% | 78.3% | 98.2% | 99.4% | 98.8% | 99.0% |
| Facts Only | 86.4% | 87.4% | 96.2% | 95.8% | 99.9% | 100.0% |
| All | 85.4% | 85.2% | 97.3% | 96.7% | 99.4% | 99.8% |

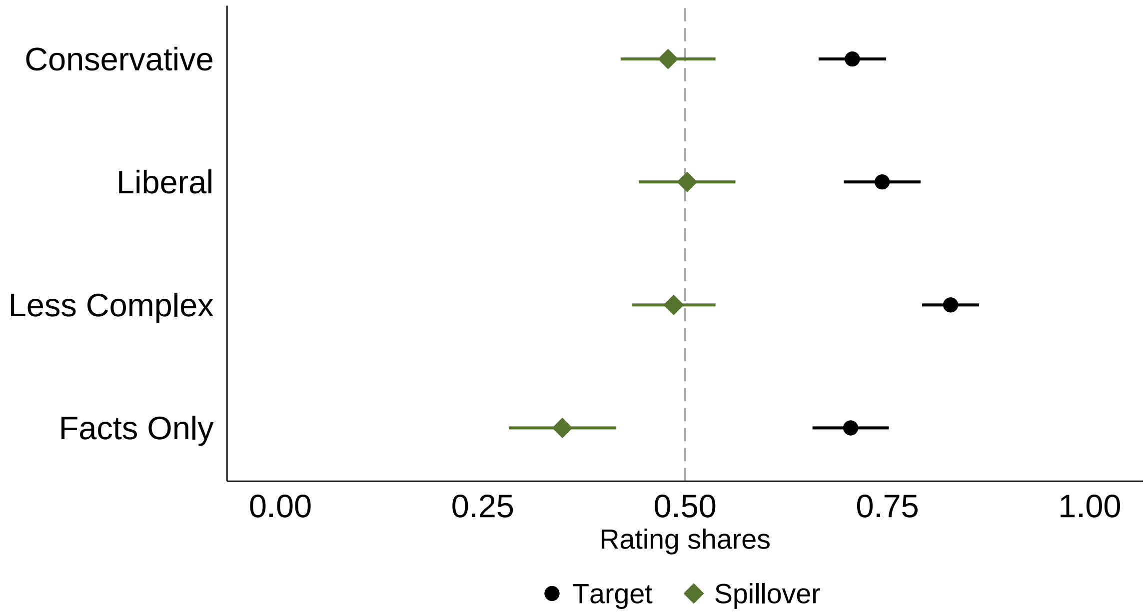
*Note:* R=Regular, V=Version. "R→V" is the share of *Regular* core facts covered by the rewrite (1–omissions). "V→R" is the share of *Version* core facts covered by *Regular* (1–hallucinations). "Embeddings" uses `all-mpnet-base-v2` at the calibrated cosine similarity threshold ($\tau$=0.66). "+ LLM Adjudication" checks unmatched pairs with an LLM to rule out false negatives. "Numeric Alignment" reports consistency of numeric content under the current repo analysis pipeline.

### B.2.2 Rewrite Validation

Fact preservation alone is not sufficient for interpreting the experiment. The rewrites must also generate framing differences that are noticeable to readers and move in the intended direction. To assess this, we conducted a human validation study in which respondents compare the *Regular* version of an article to one rewritten version and indicate which text is more conservative, more liberal, less complex, or more fact-driven.

The key advantage of this exercise is that it tests whether the intended treatment dimension is perceptible without revealing the labels of the versions to respondents. Figure B.2 reports the share of respondents who classify each rewrite along the different dimensions. The black dots show correct identification of the target dimension, while the green squares show spillovers into other dimensions. This presentation therefore allows us to assess both whether the treatment is salient and whether it unintentionally shifts perceptions along dimensions that are not meant to vary.

Figure B.2: Human Validation of Article Rewrites




*Note:* This figure shows the share of respondents rating the rewritten version of the article as being more conservative, more liberal, less complex, or more fact-driven compared to the regular version of the article in a blinded binary comparison. The black dots show the fraction of respondents correctly identifying the article version that was varied along the target dimension (e.g., *Liberal* rewrites being rated as more liberal). The green squares show the fraction of respondents classifying an article as belonging to a spillover dimension (e.g., non-*Liberal* rewrites being rated as more liberal). Error bars represent 95% confidence intervals.

The figure shows that respondents identify the intended dimension at high rates for the versions most relevant here. In particular, the *Liberal* and *Conservative* rewrites are reliably perceived as more liberal and more conservative, respectively. Respondents also perceive *Facts Only* as more fact-driven and *Less Complex* as less complex. Spillovers into other dimensions are modest overall. This pattern supports the interpretation that the political rewrites generate meaningful ideological framing differences, while the neutral versions used in the control group are perceived primarily along their intended non-political dimensions rather than as latent political treatments.

## B.3 Classification of Spin

*Prompt*

> Our goal is to determine how biased a newspaper article is towards a left or right-wing point of view, with the Democratic Party broadly representing the left and the Republican Party broadly representing the right. We will use a 7-point scale to measure the degree of bias, ranging from extreme left to extreme right.
>
> There are two main factors that we consider when assessing the bias of an article: language

and political position.

(1) When we look at the language of an article, we consider whether it uses words and phrases that are typically used by members of the Republican or Democratic parties. If an article uses more Republican-like language, we consider it biased towards the right. If it uses more Democratic-like language, we consider it biased towards the left. For example, terms like "death tax" would be considered slanted towards the right, whereas "estate tax" or "inheritance tax" would be considered slanted towards the left.

(2) Next, we look at the political position presented in the article. If the article aligns with the political views of the Republican party, we consider it biased towards the right. If it aligns with the political views of the Democratic party, we consider it biased towards the left. As an example, an article that presents an anti-abortion stance would be viewed as having a right-leaning bias, while an article that presents a pro-abortion stance would be considered to have a left-leaning bias.

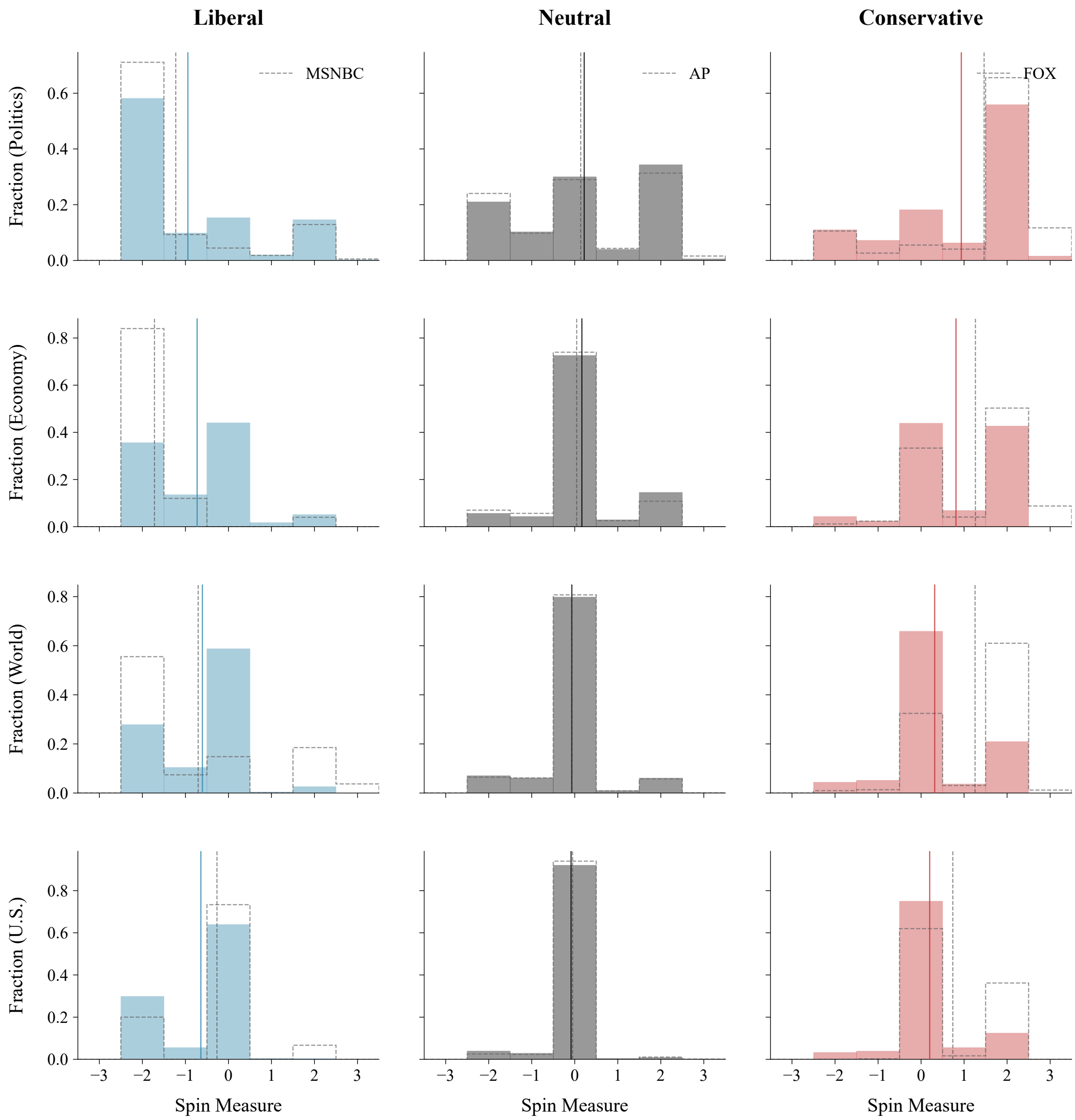
By considering these factors, we can assign a score to each article that reflects its degree of bias towards the left or right. To help you understand the spin rating scale, we've provided rankings for a few well-known politicians (we did not come up with those rankings; we took them from political scientists Keith Poole and Howard Rosenthal). You can find the rankings in the table below.

| | | |
|---|---|---|
| 1 | Very left-wing | Alexandra Ocasio-Cortez |
| 2 | Left-wing | Cory Booker |
| 3 | Somewhat left-wing | Chuck Schumer |
| 4 | Neutral | |
| 5 | Somewhat right-wing | Mitt Romney |
| 6 | Right-wing | Marco Rubio |
| 7 | Very right-wing | Ted Cruz |

You must respond with valid JSON containing exactly two fields: "score" (integer from 1-7) and "reasoning" (one sentence explanation).

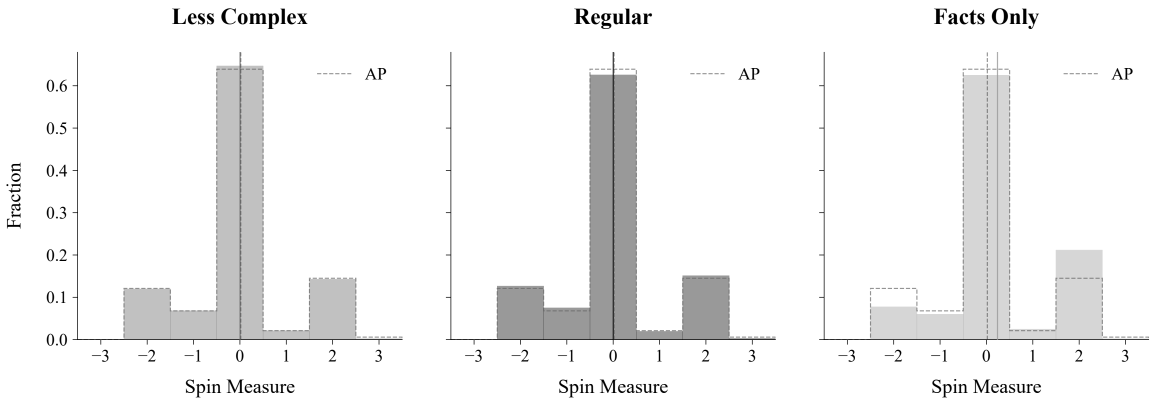
## B.4 Spin Analysis

Figure B.3: Distribution of Spin Across News Categories




*Note:* This figure shows the distribution of spin for all articles within each relevant category for the *liberal*, *politically neutral*, and *conservative* versions. The scale goes from -3 (extremely favorable to the Democratic party) to +3 (extremely favorable to the Republican party). The spin measure is based on the methodology of Braghieri et al. (2024). The dashed lines show spin for the news outlets MSNBC, Associated Press and Fox using the same measure for the same time frame and categories (standardized by the distribution of categories of our sample). The vertical lines show the means of the distributions.

Figure B.4: Distribution of Spin Within Politically Neutral Versions




*Note:* This figure shows the distribution of spin for all articles in the politically neutral versions: *regular, less complex, facts only*. The scale goes from -3 (extremely favorable to the Democratic party) to +3 (extremely favorable to the Republican party). The spin measure is based on the methodology of Braghieri et al. (2024). The dashed lines show spin for the original Associated Press articles using the same measure for the same time frame and categories. The vertical lines show the means of the distributions.

Table B.2: Drivers of Spin Including Phrases

| **Category** | **Share (%)** |
|---|---|
| **Conservative spin** | |
| **Criticism and contestation:** *Emphasized*: raising questions; sparked debate; critics claim. *Downplayed*: trump controversial; critics argue | 24.0 |
| **Fiscal and taxpayer:** *Emphasized*: taxpayer money; taxpayer funded; fiscal responsibility; use taxpayer; government overreach | 20.7 |
| **Social and moral values:** *Emphasized*: conservative values; pro life; traditional values; personal responsibility; pro abortion | 10.1 |
| **Security and order:** *Emphasized*: national security; election integrity; law and order; border security; illegal immigration | 8.3 |
| **Partisan positioning:** *Emphasized*: decisive action; liberal policies; left leaning; conservative principles. *Downplayed*: far right | 8.3 |
| **Equity and rights:** *Downplayed*: human rights; abortion rights; reproductive rights; civil rights; rights advocates | 7.2 |
| **Climate and energy:** *Downplayed*: climate change; clean energy; climate crisis; renewable energy; climate action | 5.7 |
| **National interest:** *Emphasized*: protecting american; american industries; energy independence; national sovereignty; american interests | 3.1 |
| **Markets and growth:** *Emphasized*: private sector; free market; pro growth; streamline government; streamline federal | 2.5 |

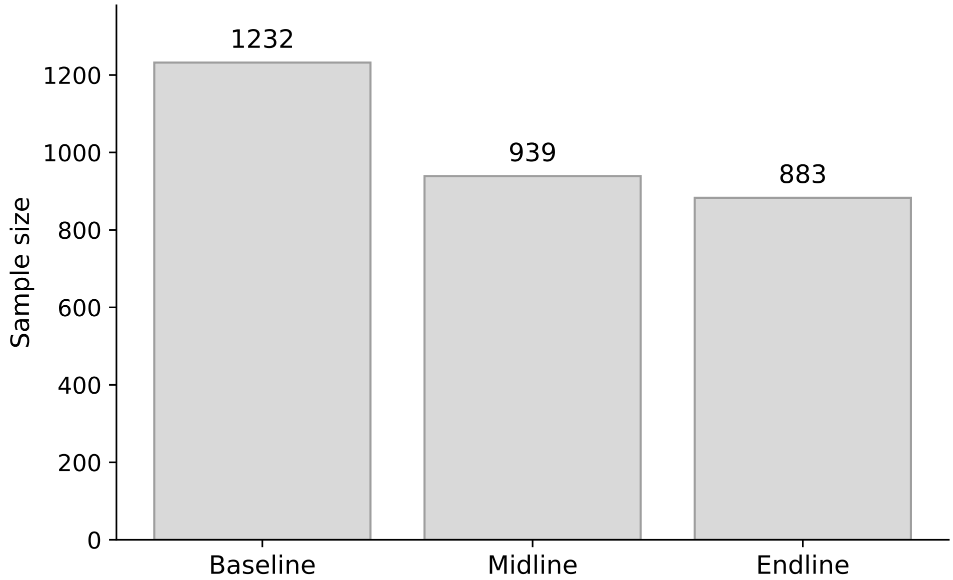
Table B.2: Drivers of Spin Including Phrases (continued)

| Category | Share (%) |
|---|---|
| **Liberal spin** | |
| **Urgency and crisis:** *Emphasized*: urgent need; critical need; humanitarian crisis; critical issues; urgent need comprehensive | 25.8 |
| **Political actors:** *Emphasized*: trump administration; trump divisive; president trump controversial. *Downplayed*: president trump; president donald trump | 22.9 |
| **Emphasis and salience:** *Emphasized*: highlighting urgent; highlighting urgent need; underscores urgent; underscoring urgent need; highlighting need | 19.2 |
| **Public investment and services:** *Emphasized*: public investments; public health; public services; essential public services | 10.1 |
| **Equity and rights:** *Emphasized*: human rights; reproductive rights; systemic issues; highlighting systemic; address systemic | 5.1 |
| **Criticism and contestation:** *Emphasized*: raising concerns; trump controversial; critics argue; raising alarms. *Downplayed*: expressed concerns | 3.2 |
| **Security and order:** *Downplayed*: law enforcement; national security; security concerns; security measures; border security | 2.9 |
| **Institutions and accountability:** *Emphasized*: democratic values; international cooperation; democratic process; corporate accountability; democratic norms | 1.5 |

*Note:* This table decomposes the mean change in the spin index induced by political rewrites into phrasing categories, shown separately for conservative and liberal rewrites relative to the corresponding regular version. Categories are constructed by grouping short phrase families (2–3 word n-grams), and category contributions are obtained by aggregating phrase-level contributions to the mean spin shift. Listed under each category are the top five positively contributing raw phrases, labeled as *Emphasized* or *Downplayed* depending on whether the phrase is used more or less in the target rewrite relative to the regular version. Shares are percentages of total positive category contribution on each side. Only categories with share $\geq 1\%$ are shown.

## B.5 Summary Statistics

Figure B.5: Sample size




*Note:* This figure shows the sample size at baseline, at midline, and at endline.

Table B.3: Summary Statistics

| | Mean | SD | Min | Max |
|---|---|---|---|---|
| **Demographics** | | | | |
| Age | 37.267 | 11.950 | 18 | 76 |
| Female | 0.571 | 0.495 | 0 | 1 |
| White | 0.683 | 0.466 | 0 | 1 |
| At least some college | 0.869 | 0.337 | 0 | 1 |
| **Political leaning** | | | | |
| Democrat | 0.429 | 0.495 | 0 | 1 |
| Independent | 0.332 | 0.471 | 0 | 1 |
| Republican | 0.239 | 0.426 | 0 | 1 |
| **News consumption** | | | | |
| News primarily online | 0.903 | 0.297 | 0 | 1 |
| Somewhat frequent news | 0.585 | 0.493 | 0 | 1 |
| News apps | 1.843 | 1.654 | 0 | 6 |
| Observations | 1232 | | | |

*Note:* This table presents summary statistics of the sample. "Age" is the respondent age in years. "Female", "White", and "At least some college" are binary indicators. The political leaning measures are binary indicators for being Democrat, Independent, or Republican. The news consumption measure refers to the share of respondents indicating that they primarily read news online ("News primarily online") and read the news at least somewhat frequently ("Somewhat frequent news"), and "News apps" indicates the number of news apps on the phone.

Table B.4: Balance Table

| | (1) Control | (2) Aligned | (3) Misaligned | (4) p-val (2) vs (1) | (5) p-val (3) vs (1) |
|---|---|---|---|---|---|
| **Demographics** | | | | | |
| Age | 37.444 | 37.375 | 36.608 | 0.934 | 0.339 |
| | (12.315) | (11.203) | (11.581) | | |
| Female | 0.553 | 0.610 | 0.588 | 0.117 | 0.352 |
| | (0.497) | (0.489) | (0.493) | | |
| White | 0.703 | 0.689 | 0.613 | 0.681 | 0.012 |
| | (0.457) | (0.464) | (0.488) | | |
| Some college | 0.869 | 0.853 | 0.887 | 0.520 | 0.442 |
| | (0.338) | (0.355) | (0.317) | | |
| **Political leaning** | | | | | |
| Democrat | 0.418 | 0.438 | 0.454 | 0.583 | 0.333 |
| | (0.494) | (0.497) | (0.499) | | |
| Independent | 0.351 | 0.287 | 0.321 | 0.057 | 0.390 |
| | (0.478) | (0.453) | (0.468) | | |
| Republican | 0.231 | 0.275 | 0.225 | 0.171 | 0.853 |
| | (0.422) | (0.447) | (0.418) | | |
| **News consumption** | | | | | |
| News primarily online | 0.896 | 0.916 | 0.908 | 0.331 | 0.574 |
| | (0.305) | (0.277) | (0.289) | | |
| Somewhat frequent news | 0.579 | 0.614 | 0.575 | 0.334 | 0.915 |
| | (0.494) | (0.488) | (0.495) | | |
| News apps | 1.823 | 1.801 | 1.950 | 0.853 | 0.308 |
| | (1.641) | (1.666) | (1.684) | | |
| Observations | 741 | 251 | 240 | | |

*Note:* This table presents summary statistics by treatment group. "Age" is the respondent age in years. "Female", "White", and "At least some college" are binary indicators. The political leaning measures are binary indicators for being Democrat, Independent, or Republican. The news consumption measure refers to the share of respondents indicating that they primarily read news online (“News primarily online”) and read the news at least somewhat frequently (“Somewhat frequent news”), and “News apps” indicates the number of news apps on the phone. Columns (1), (2), and (3) show means with standard errors in parentheses for the control group, the aligned group, and the misaligned group. Column (4) reports p-values for tests of equality between the aligned and the control group. Column (5) reports p-values for tests of equality between the misaligned and the control group.

Table B.5: Attrition

| | (1) Midline | (2) Endline |
|---|---|---|
| Aligned | 0.0193 | 0.000899 |
| | (0.0316) | (0.0340) |
| Misaligned | 0.0200 | 0.0409 |
| | (0.0316) | (0.0329) |
| Constant | 0.765*** | 0.677*** |
| | (0.0259) | (0.0286) |
| Wave FE | Yes | Yes |
| Party control | Yes | Yes |
| N | 1232 | 1232 |

*Note:* This table reports the results of regressions of an indicator for participation in the midline or endline survey on treatment assignment. Each coefficient reflects the difference in participation probability relative to the control group. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table B.6: Baseline Predictors of Reported Willingness to Download the App

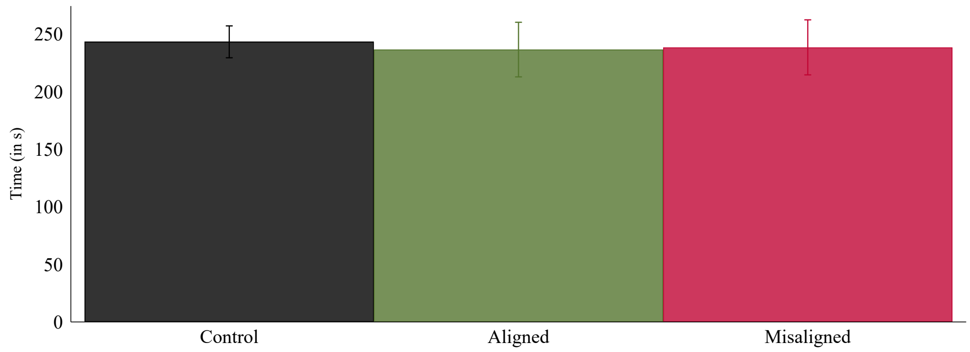
| | (1)<br>App Download Willingness |
|---|---|
| Female | 0.00303 |
| | (0.0100) |
| Age | 0.000612 |
| | (0.000458) |
| White | 0.0145 |
| | (0.0123) |
| At least some college | -0.0355*** |
| | (0.00978) |
| Full-time employed | 0.000414 |
| | (0.0103) |
| Income at least $75,000 | 0.00248 |
| | (0.0112) |
| News primarily online | 0.00711 |
| | (0.0186) |
| News primarily on phone | -0.00345 |
| | (0.0119) |
| Number of news apps | -0.00661 |
| | (0.00419) |
| At least somewhat frequent news use | 0.00794 |
| | (0.0116) |
| Trust in AI | -0.00682 |
| | (0.00675) |
| Outcome mean | 0.966 |
| Wave FE | Yes |
| N | 1232 |

*Note:* This table reports a regression of reported willingness to download the app on baseline respondent characteristics. “Willingness” refers to respondents’ reported willingness to download the app and should not be interpreted as a direct measure of realized download completion. The specification includes wave fixed effects. The bottom rows report the outcome mean, whether wave fixed effects are included, and the sample size. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

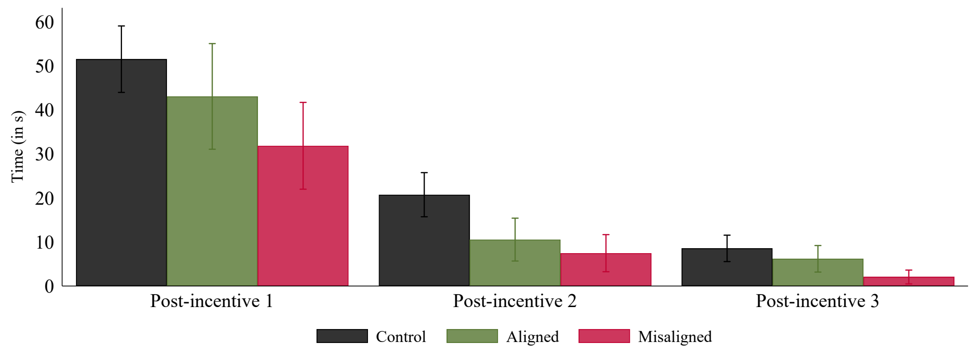
## B.6 Raw Consumption Data

Figure B.6: Raw Consumption Data: Time

(a) **Incentive period**



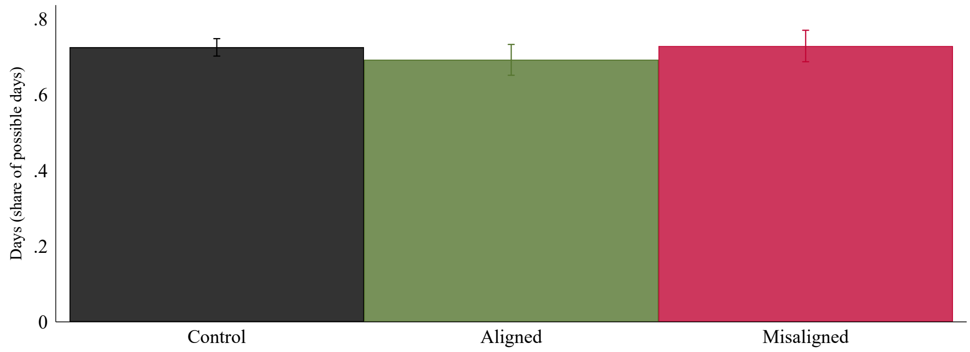

(b) **Post-incentive period**



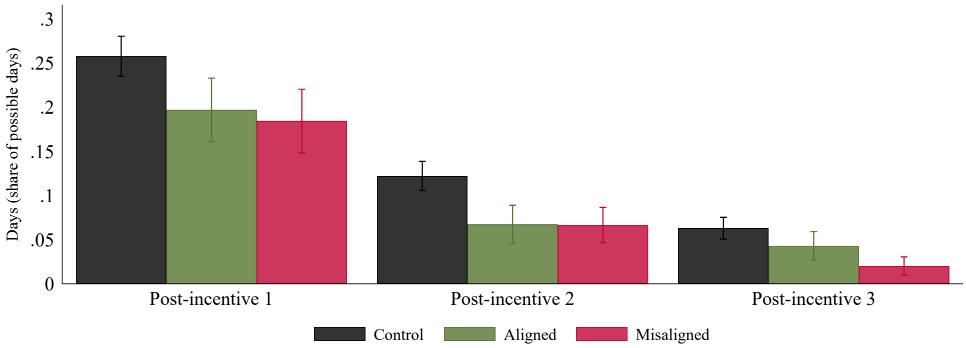

*Note:* This figure shows the raw data on winsorized daily time spent in the app by treatment group. Panel (a) presents the incentive period; Panel (b) presents the post-incentive period. The post-incentive period is divided into 10-day intervals. Error bars represent 95% confidence intervals.

Figure B.7: Raw Consumption Data: Days

(a) **Incentive period**



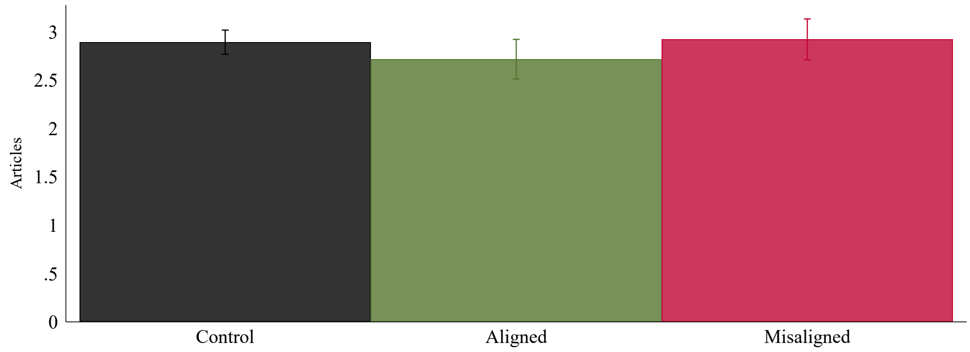

(b) **Post-incentive period**



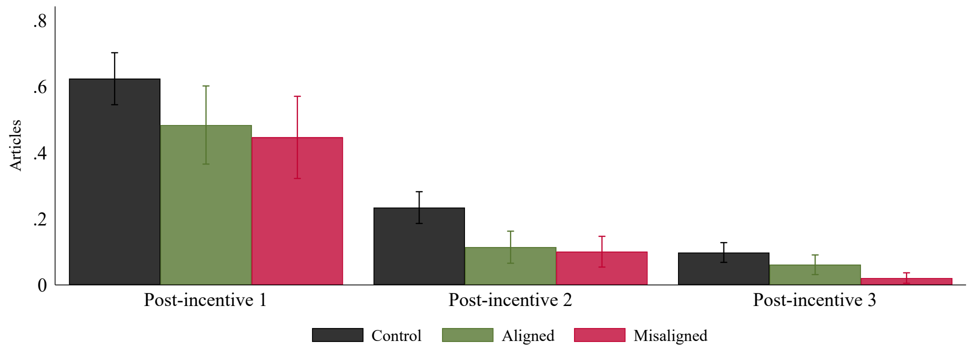

*Note:* This figure shows the raw data on share of days using the app by treatment group. Panel (a) presents the incentive period; Panel (b) presents the post-incentive period. The post-incentive period is divided into 10-day intervals. Error bars represent 95% confidence intervals.

Figure B.8: Raw Consumption Data: Articles

(a) **Incentive period**




(b) **Post-incentive period**

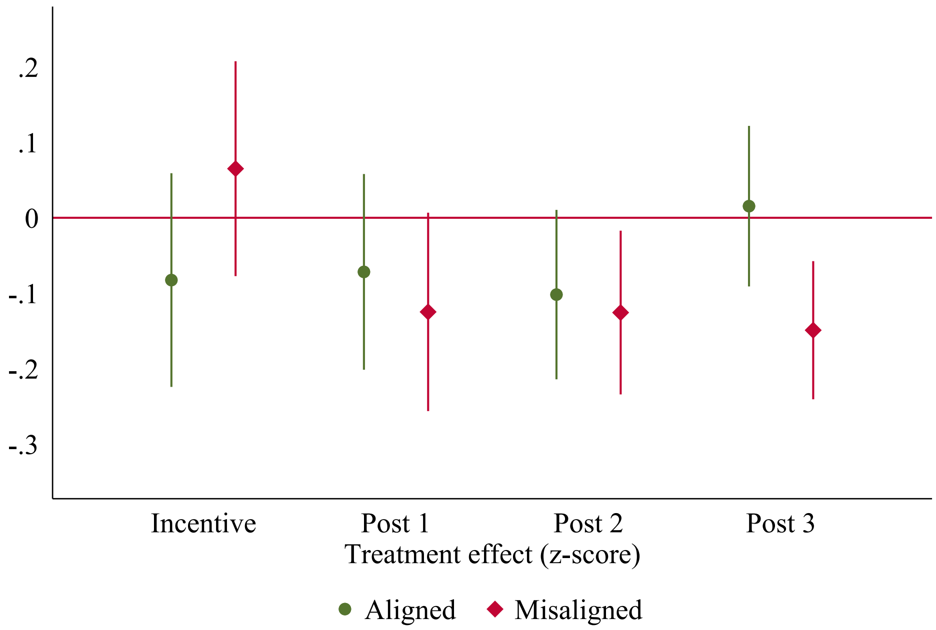



*Note:* This figure shows the raw data on the number of daily articles read in the app by treatment group. Panel (a) presents the incentive period; Panel (b) presents the post-incentive period. The post-incentive period is divided into 10-day intervals. Error bars represent 95% confidence intervals.

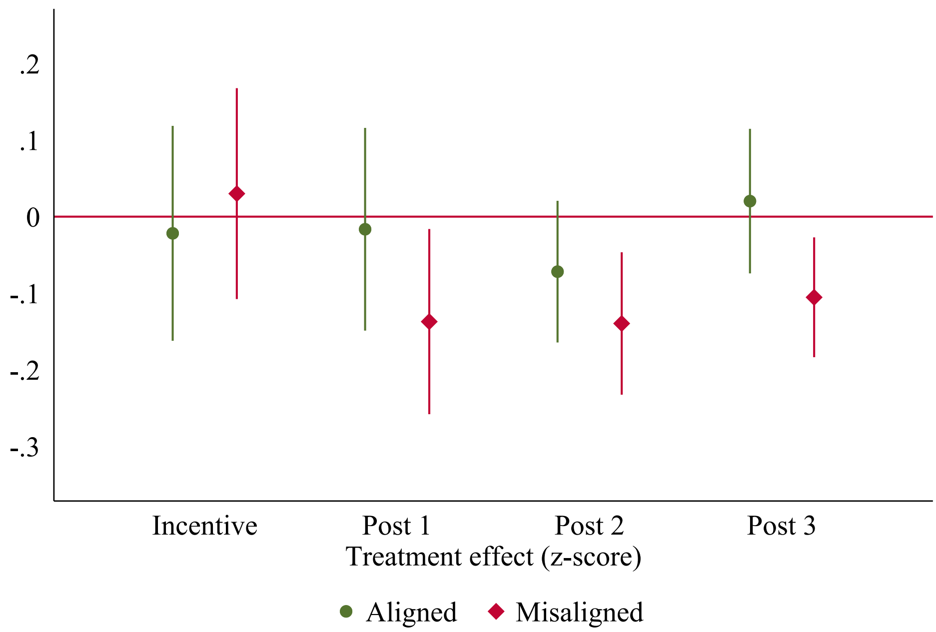
## B.7 Consumption by Period

Figure B.9: Effects on News Consumption by Period

(a) **Days**




(b) **Time**




*Note:* This figure shows the effect of aligned and misaligned versions on news consumption over time. Incentive period refers to the 10 days in which users were paid to use the app. Post-incentive periods pool 10 days each. Panel (a) presents the share of days using the app; Panel (b) presents winsorized average daily time spent using the app. Both outcomes are standardized. Error bars represent 95% confidence intervals.

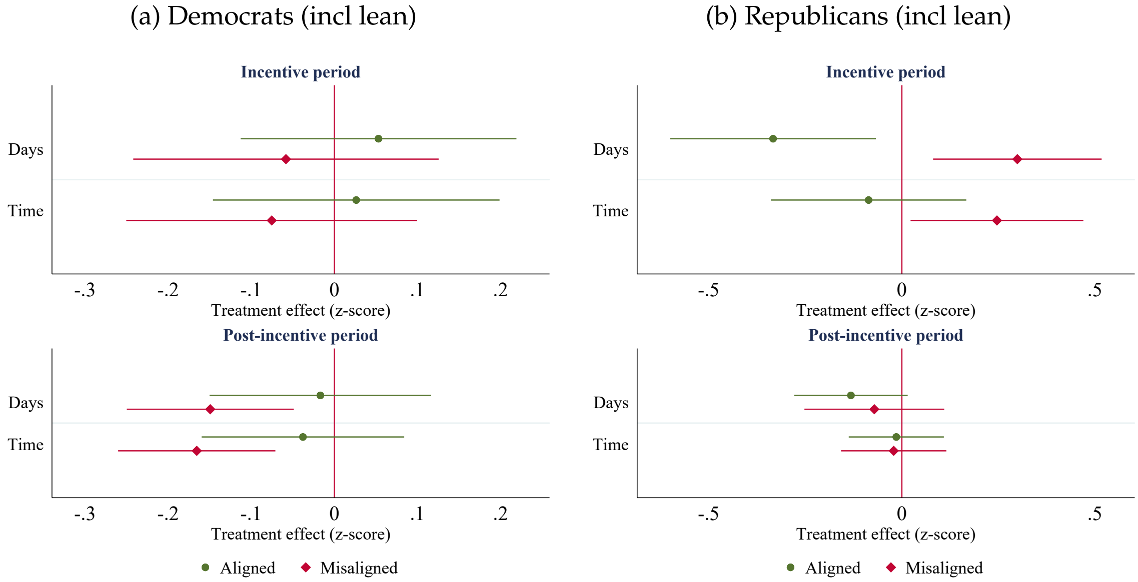
## B.8 Exposure Validation During the Incentive Period

Table B.7: Incentive-Period Exposure to Focal News Categories

| | (1) Article opens | (2) Time (s) | (3) Open share | (4) Time share |
|---|---|---|---|---|
| Aligned | -0.336** | -7.523 | 0.0150 | -0.0254* |
| | (0.131) | (16.97) | (0.0163) | (0.0147) |
| Misaligned | -0.0569 | -7.331 | 0.000854 | -0.0321** |
| | (0.139) | (15.53) | (0.0176) | (0.0150) |
| Control mean | 2.584 | 204.083 | 0.728 | 0.866 |
| P-value | 0.067 | 0.992 | 0.488 | 0.721 |
| Controls | Yes | Yes | Yes | Yes |
| Wave FE | Yes | Yes | Yes | Yes |
| N | 1232 | 1232 | 1158 | 1129 |

*Note:* This table reports incentive-period regressions of focal-category exposure on aligned and misaligned treatment assignment, relative to the neutral control group. The focal categories are *Politics*, *Economy*, *U.S.*, and *World*. "Article opens" and "Time (s)" are period-level mean daily exposure measures. "Open share" and "Time share" are the shares of total article opens and category-level reading time accounted for by the focal categories. Share outcomes are defined for respondents with positive total article opens or category-level reading time in the incentive period. All specifications include the same baseline covariates, party controls, wave fixed effects, and onboarding-date fixed effects as the main consumption regressions. Standard errors clustered at the respondent level in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

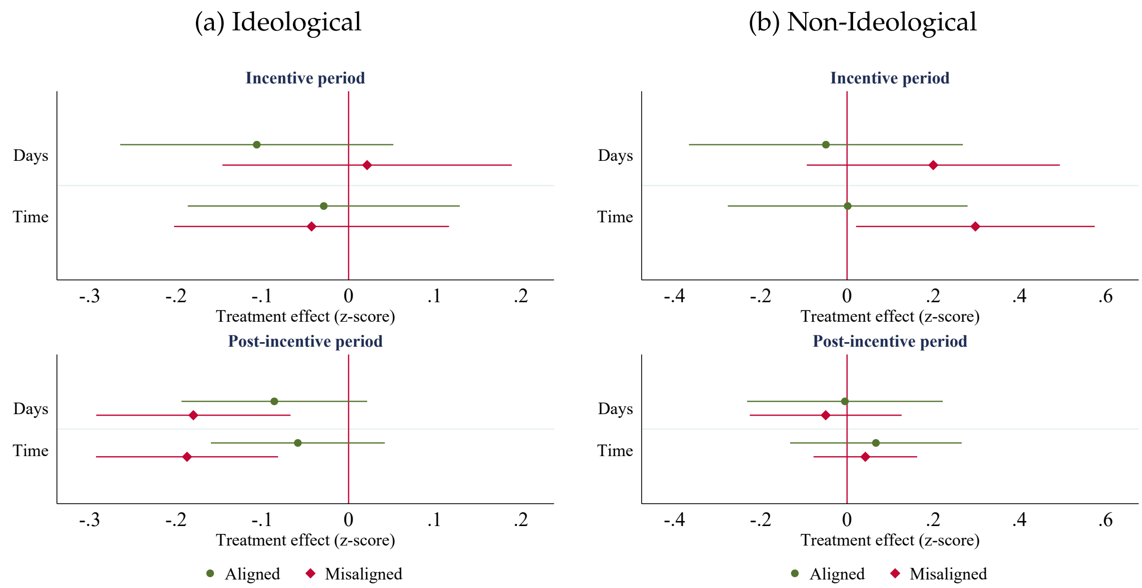
Figure B.10: Effects on News Consumption by Democrats and Republicans




*Note:* This figure shows the effect of aligned and misaligned versions on news consumption in the incentive period (top panel) and post-incentive period (bottom panel). Panel (a) presents effects for Democrats; Panel (b) presents effects for Republicans. News consumption is measured as the share of days using the app and winsorized daily time spent using the app. Both outcomes are standardized. Error bars represent 95% confidence intervals.

## B.9 Political Heterogeneity

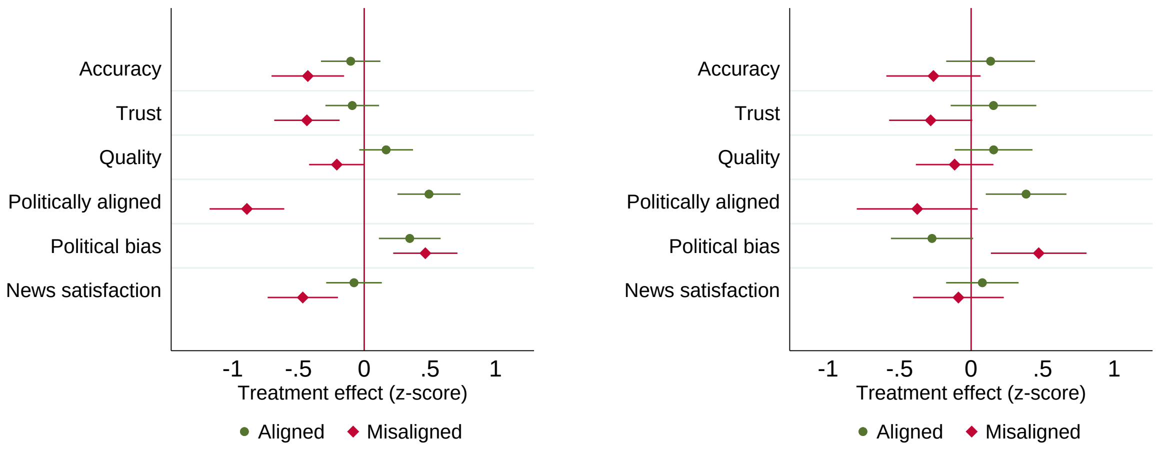
Figure B.11: Effects on News Consumption by Ideology




*Note:* This figure shows the effect of aligned and misaligned versions on news consumption in the incentive period (top panel) and post-incentive period (bottom panel). Panel (a) presents effects for ideological respondents; Panel (b) presents effects for non-ideological respondents. Ideological refers to identifying as either liberal or conservative, and non-ideological refers to identifying as "neither conservative nor liberal." News consumption is measured as the share of days using the app and winsorized daily time spent using the app. Both outcomes are standardized. Error bars represent 95% confidence intervals.

Figure B.12: Effects on Perceptions and News Satisfaction by Democrats and Republicans

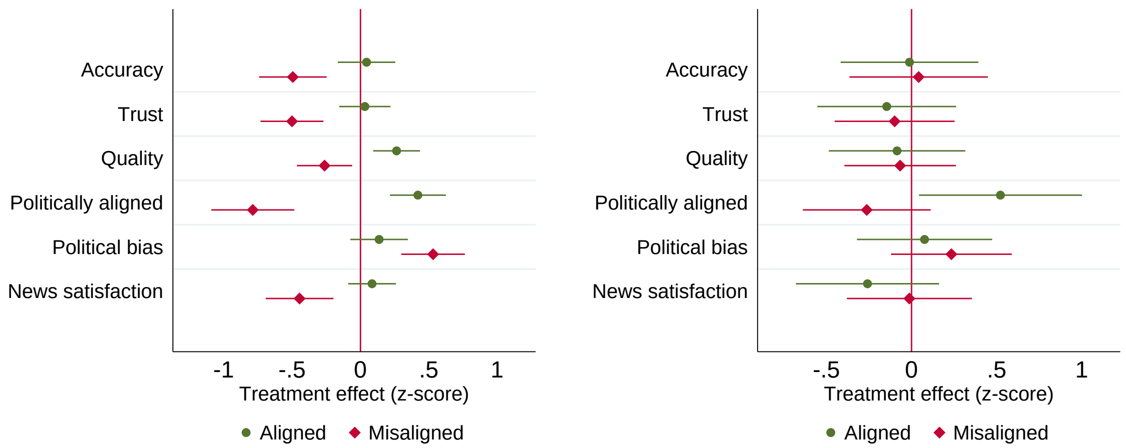
(a) Democrats (incl lean) (b) Republicans (incl lean)




*Note:* This figure shows the effect of aligned and misaligned versions on perceptions of the app and news satisfaction. Panel (a) presents effects for Democrats; Panel (b) presents effects for Republicans. *Accuracy*, *Trust*, and *Quality* measure respondent ratings of the articles in the app. Political ratings are based on respondent ratings of articles on a 5-point scale from "very right-wing biased" to "very left-wing biased." *Politically aligned* is defined as proximity to "very right-wing biased" for Republicans and "very left-wing biased" for Democrats. *Political bias* is an indicator for rating it as biased in any direction (as compared to "not biased"). *News satisfaction* is respondents' reported satisfaction with the news they have been reading over the past week. All outcomes are standardized. Error bars represent 95% confidence intervals.

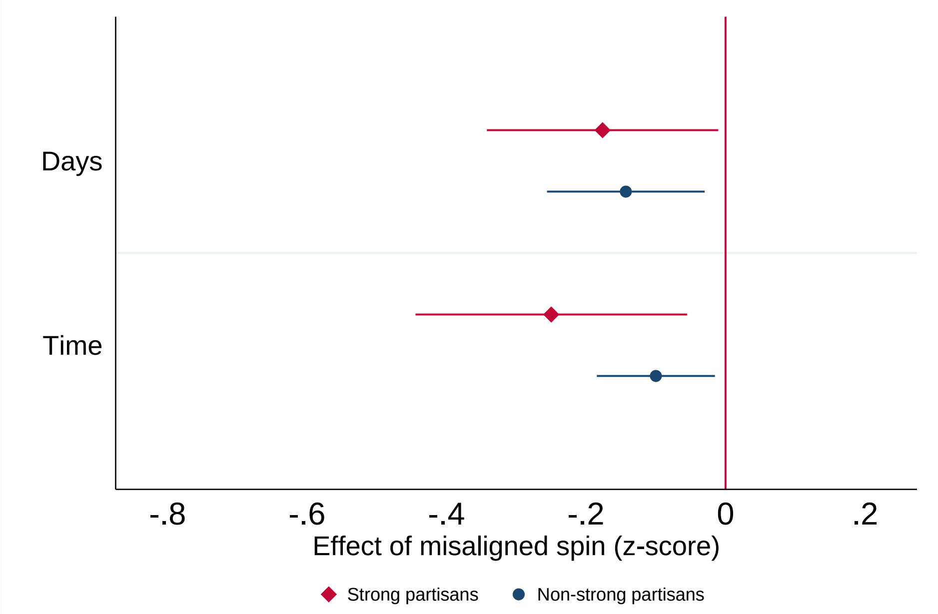
Figure B.13: Effects on Perceptions and News Satisfaction by Ideology

(a) Ideological (b) Non-Ideological



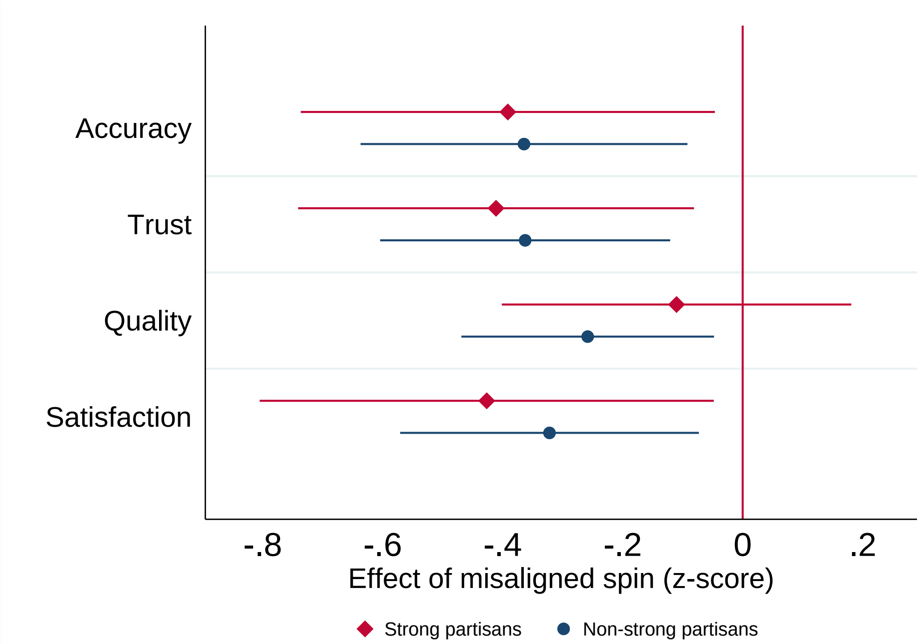

*Note:* This figure shows the effect of aligned and misaligned versions on perceptions of the app and news satisfaction. Panel (a) presents effects for ideological respondents; Panel (b) presents effects for non-ideological respondents. *Accuracy*, *Trust*, and *Quality* measure respondent ratings of the articles in the app. Political ratings are based on respondent ratings of articles on a 5-point scale from "very right-wing biased" to "very left-wing biased". *Politically aligned* is defined as proximity to "very right-wing biased" for Republicans and "very left-wing biased" for Democrats. *Political bias* is an indicator for rating it as biased in any direction (as compared to "not biased"). *News satisfaction* is respondents' reported satisfaction with the news they have been reading over the past week. All outcomes are standardized. Error bars represent 95% confidence intervals.

Figure B.14: Effects of Misaligned Spin on Post-Incentive Consumption by Strength of Partisanship




*Note:* This figure shows the effect of *misaligned* spin on post-incentive news consumption by strength of partisanship. Outcomes are the share of days using the app and winsorized daily time spent in the app. "Strong partisans" are respondents who identify as strong Democrats or strong Republicans. "Non-strong partisans" are respondents who identify as weak partisans or independents leaning toward one party. All outcomes are standardized. Error bars represent 95% confidence intervals.

Figure B.15: Effects of Misaligned Spin on Perceptions and Satisfaction by Strength of Partisanship

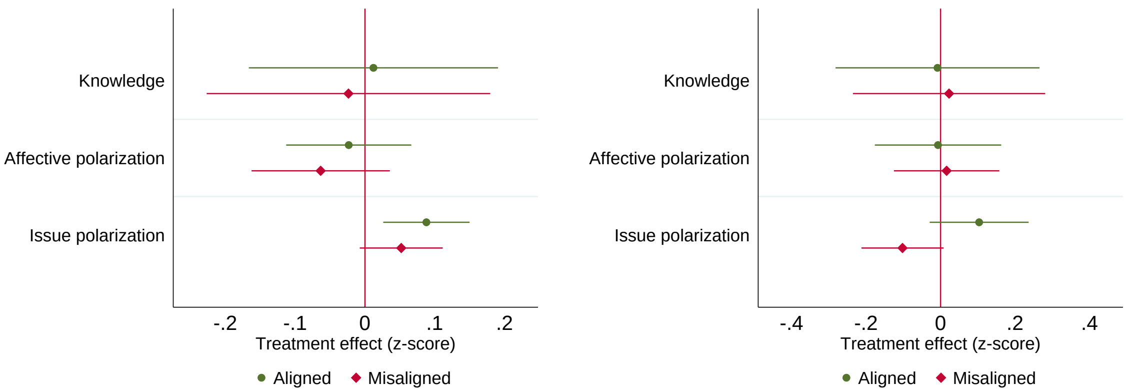



*Note:* This figure shows the effect of *misaligned* spin on midline perceptions and news satisfaction by strength of partisanship. Outcomes include perceived accuracy, trust, quality, and overall news satisfaction. "Strong partisans" are respondents who identify as strong Democrats or strong Republicans. "Non-strong partisans" are respondents who identify as weak partisans or independents leaning toward one party. All outcomes are standardized. Error bars represent 95% confidence intervals.

Figure B.16: Effects on Knowledge and Polarization by Democrats and Republicans

(a) Democrats (incl lean)

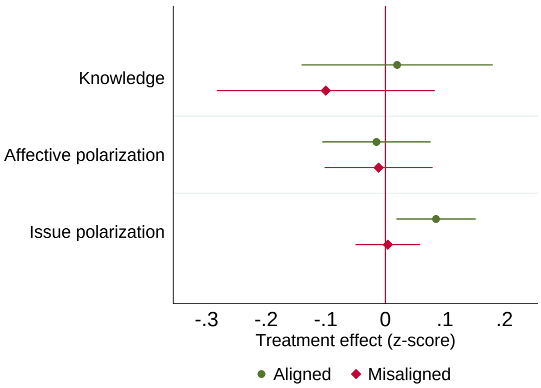
(b) Republicans (incl lean)



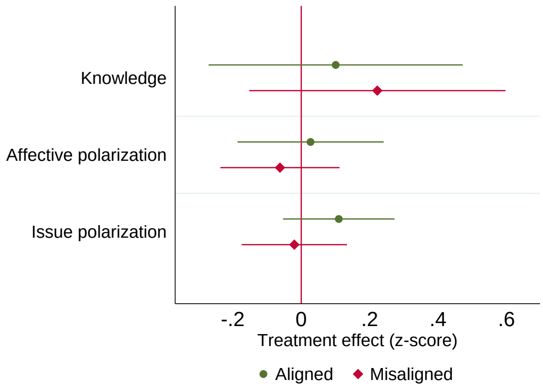

*Note:* This figure shows the effect of aligned and misaligned versions on knowledge and polarization. Panel (a) presents effects for Democrats; Panel (b) presents effects for Republicans. *Knowledge* refers to the score in a quiz of 15 questions based on recent news stories. *Affective polarization* is an index combining feeling warmly toward Republicans and Democrats and hypothetical division of $100 between a Democrat and Republican voter. *Issue polarization* is an index combining policy views on immigration, firearms, international trade, racial discrimination by police, and sexual harassment. All outcomes are standardized. Error bars represent 95% confidence intervals.

Figure B.17: Effects on Knowledge and Polarization by Ideology

(a) Ideological

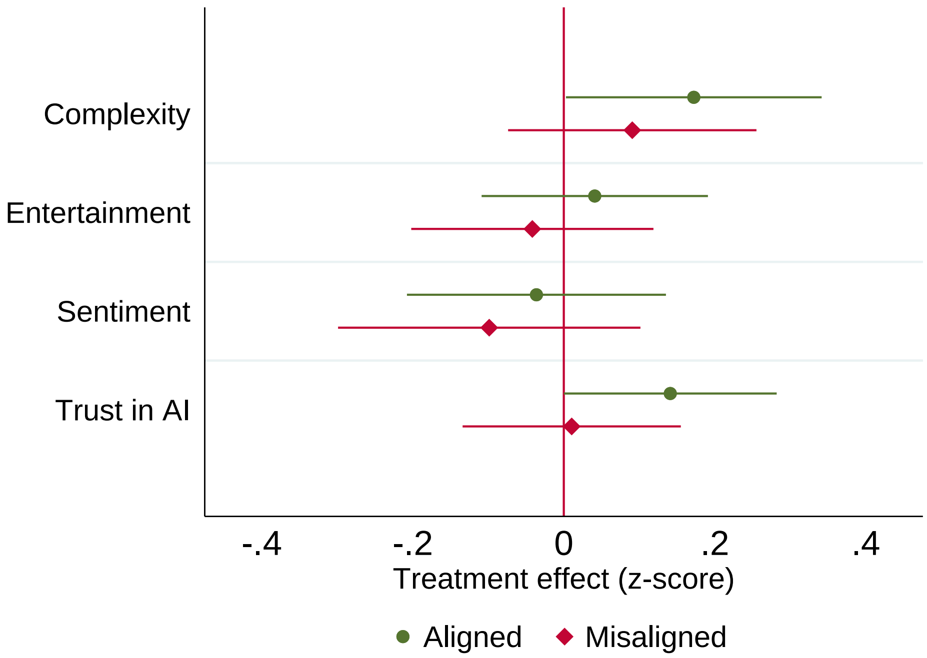



(b) Non-Ideological




*Note:* This figure shows the effect of aligned and misaligned versions on knowledge and polarization. Panel (a) presents effects for ideological respondents; Panel (b) presents effects for non-ideological respondents. *Knowledge* refers to the score in a quiz of 15 questions based on recent news stories. *Affective polarization* is an index combining feeling warmly toward Republicans and Democrats and hypothetical division of $100 between a Democrat and Republican voter. *Issue polarization* is an index combining policy views on immigration, firearms, international trade, racial discrimination by police, and sexual harassment. All outcomes are standardized. Error bars represent 95% confidence intervals.

Figure B.18: Effects on Perceptions of Complexity, Entertainment, Sentiment and Trust in AI




*Note:* This figure shows the effect of aligned and misaligned versions on perceptions of the app and trust in AI. *Complexity*, *Entertainment*, and *Sentiment* refer to respondent ratings of the articles in the app. *Trust in AI* measures respondents' general trust in AI on a 5-point scale. All outcomes are standardized. Error bars represent 95% confidence intervals.

## B.10 Additional Tables

Table B.8: Effects on News Consumption

| | Incentive | | Post-incentive | |
|---|---|---|---|---|
| | (1) Days | (2) Time | (3) Days | (4) Time |
| Aligned | -0.0826 | -0.0218 | -0.0527 | -0.0226 |
| | (0.0722) | (0.0715) | (0.0514) | (0.0464) |
| Misaligned | 0.0651 | 0.0301 | -0.133*** | -0.127*** |
| | (0.0726) | (0.0702) | (0.0464) | (0.0414) |
| P-value | 0.096 | 0.547 | 0.149 | 0.025 |
| Controls | Yes | Yes | Yes | Yes |
| Wave FE | Yes | Yes | Yes | Yes |
| N | 1232 | 1232 | 3696 | 3696 |

*Note:* This table shows the effect of aligned and misaligned versions on news consumption during the incentive period and the pooled post-incentive period. News consumption is measured as the share of days using the app and winsorized daily time spent using the app. Both outcomes are standardized. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table B.9: Effects on Perceptions and News Satisfaction

| | (1) Accuracy | (2) Trust | (3) Quality | (4) Politically aligned | (5) Political Bias | (6) News satisfaction |
|---|---|---|---|---|---|---|
| Aligned | -0.00315 | -0.0226 | 0.167** | 0.418*** | 0.142 | -0.0149 |
| | (0.0914) | (0.0825) | (0.0804) | (0.0937) | (0.0910) | (0.0816) |
| Misaligned | -0.383*** | -0.392*** | -0.192** | -0.688*** | 0.473*** | -0.332*** |
| | (0.108) | (0.0979) | (0.0861) | (0.126) | (0.0996) | (0.107) |
| P-value: aligned=misaligned | 0.002 | 0.001 | 0.000 | 0.000 | 0.005 | 0.007 |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Baseline control | No | No | No | No | No | Yes |
| Wave FE | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 938 | 938 | 938 | 938 | 938 | 938 |

*Note:* This table shows the effect of aligned and misaligned versions on perceptions of the app and news satisfaction. *Accuracy*, *Trust*, and *Quality* measure respondent ratings of the articles in the app. Political ratings are based on respondent ratings of articles on a 5-point scale from "very right-wing biased" to "very left-wing biased". *Politically aligned* is defined as proximity to "very right-wing biased" for Republicans and "very left-wing biased" for Democrats. *Political bias* is an indicator for rating it as biased in any direction (as compared to "not biased"). *News satisfaction* is respondents' reported satisfaction with the news they have been reading over the past week. All outcomes are standardized. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table B.10: Midline Perceptions and Later News Consumption

| | Days | Time |
|---|---|---|
| Midline accuracy | 0.058** | 0.062*** |
| | ( 0.023) | ( 0.018) |
| Midline trust | 0.057** | 0.036 |
| | ( 0.026) | ( 0.024) |
| Midline quality | 0.059** | 0.061*** |
| | ( 0.027) | ( 0.023) |
| Midline news satisfaction | 0.056** | 0.051** |
| | ( 0.024) | ( 0.022) |
| Baseline controls | Yes | Yes |
| Treatment controls | Yes | Yes |
| Party controls | Yes | Yes |
| Wave FE | Yes | Yes |
| Onboarding-date FE | Yes | Yes |
| N | 938 | 938 |

*Note:* This table reports user-level regressions of post-incentive app use on midline perceptions of the app. The outcome variables are the average share of post-incentive days with app use and the average post-incentive time spent in the app, aggregated across periods 2–4 and standardized. The explanatory variables are the four midline perception measures shown in the rows: perceived accuracy, trust, quality, and news satisfaction. All specifications include treatment controls, party controls, baseline covariates, wave fixed effects, and onboarding-date fixed effects. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table B.11: Effects on Knowledge and Polarization

| | (1) Knowledge | (2) Affective Polarization | (3) Issue Polarization |
|---|---|---|---|
| Aligned | 0.0350 | -0.0120 | 0.0900*** |
| | (0.0737) | (0.0411) | (0.0305) |
| Misaligned | -0.0203 | -0.0381 | -0.00198 |
| | (0.0813) | (0.0402) | (0.0277) |
| P-value: aligned=misaligned | 0.563 | 0.576 | 0.011 |
| Controls | Yes | Yes | Yes |
| Baseline control | No | Yes | Yes |
| Wave FE | Yes | Yes | Yes |
| N | 938 | 938 | 938 |

*Note:* This table shows the effect of aligned and misaligned versions on knowledge and polarization. *Knowledge* refers to the score in a quiz of 15 questions based on recent news stories. *Affective polarization* is an index combining feeling warmly toward Republicans and Democrats and hypothetical division of $100 between a Democrat and Republican voter. *Issue polarization* is an index combining policy views on immigration, firearms, international trade, racial discrimination by police, and sexual harassment. All outcomes are standardized. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table B.12: Downstream Outcomes and Multiple-Testing Adjustments

| | Aligned | | | | Misaligned | | | |
|---|---|---|---|---|---|---|---|---|
| | Coef. | Raw p | Holm p | BH q | Coef. | Raw p | Holm p | BH q |
| Knowledge | 0.035 (0.074) | 0.635 | 1.000 | 0.770 | -0.020 (0.081) | 0.803 | 1.000 | 0.943 |
| Affective polarization | -0.012 (0.041) | 0.770 | 1.000 | 0.770 | -0.038 (0.040) | 0.343 | 1.000 | 0.943 |
| Issue polarization | 0.090*** (0.031) | 0.003 | 0.010 | 0.010 | -0.002 (0.028) | 0.943 | 1.000 | 0.943 |
| N | 938 | | | | 938 | | | |

*Note:* This table reports the same midline downstream treatment effects as Table B.11 and adds family-level multiple-testing adjustments across the paper's three headline downstream outcomes: knowledge, affective polarization, and issue polarization. "Holm p" reports Holm step-down adjusted p-values within the three aligned-outcome tests and within the three misaligned-outcome tests. "BH q" reports Benjamini-Hochberg q-values within those same two three-hypothesis families. Raw p-values are computed from the corresponding coefficient estimates and robust standard errors underlying Table B.11. Robust standard errors are shown in parentheses inside the coefficient cells. * p<0.1, ** p<0.05, *** p<0.01.

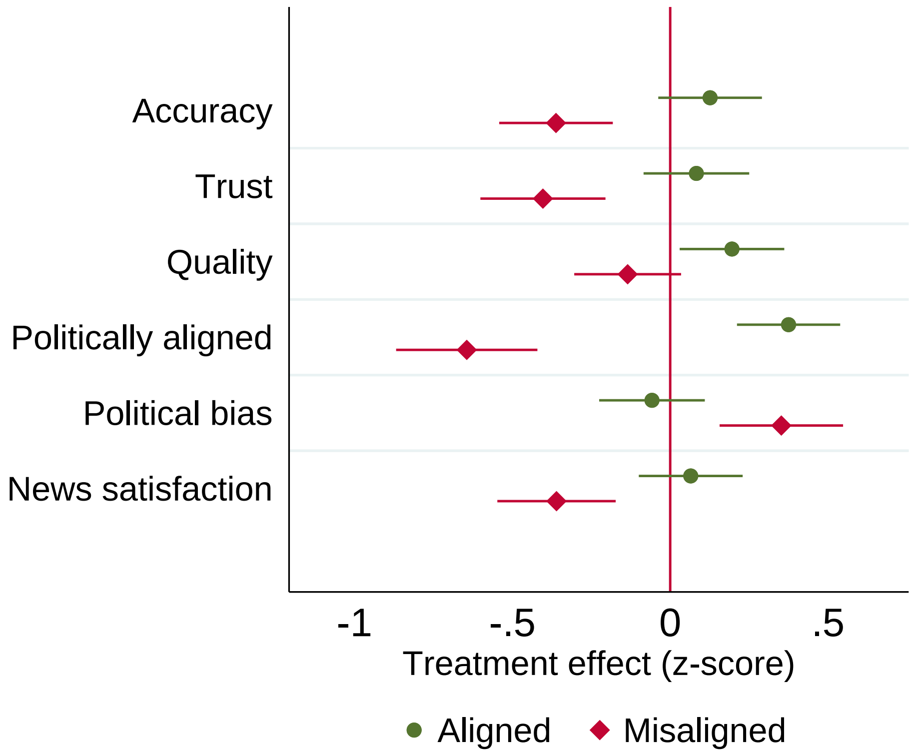
Table B.13: Issue-Polarization Components

| | Aligned | | | | Misaligned | | | |
|---|---|---|---|---|---|---|---|---|
| | Coef. | Raw p | Holm p | BH q | Coef. | Raw p | Holm p | BH q |
| Immigration | 0.083<br>(0.059) | 0.158 | 0.474 | 0.263 | -0.076<br>(0.057) | 0.188 | 0.939 | 0.435 |
| Firearms | 0.024<br>(0.044) | 0.574 | 1.000 | 0.718 | 0.048<br>(0.043) | 0.261 | 0.985 | 0.435 |
| International trade | 0.235***<br>(0.085) | 0.006 | 0.028 | 0.028 | 0.051<br>(0.077) | 0.513 | 1.000 | 0.513 |
| Racial discrimination by police | 0.001<br>(0.084) | 0.995 | 1.000 | 0.995 | -0.091<br>(0.079) | 0.246 | 0.985 | 0.435 |
| Sexual harassment | 0.148**<br>(0.073) | 0.042 | 0.168 | 0.105 | 0.048<br>(0.072) | 0.504 | 1.000 | 0.513 |
| N | 938 | | | | 938 | | | |

*Note:* This table decomposes the midline issue-polarization result into the five standardized component outcomes that enter the composite index: immigration, firearms, international trade, racial discrimination by police, and sexual harassment. Each row reports a separate regression of the component outcome on aligned and misaligned treatment assignment, controlling for the corresponding baseline component, party controls, baseline covariates, wave fixed effects, and onboarding-date fixed effects. "Holm p" reports Holm step-down adjusted p-values within the five aligned-component tests and within the five misaligned-component tests. "BH q" reports Benjamini-Hochberg q-values within those same two five-hypothesis families. Robust standard errors are shown in parentheses inside the coefficient cells. * p<0.1, ** p<0.05, *** p<0.01.

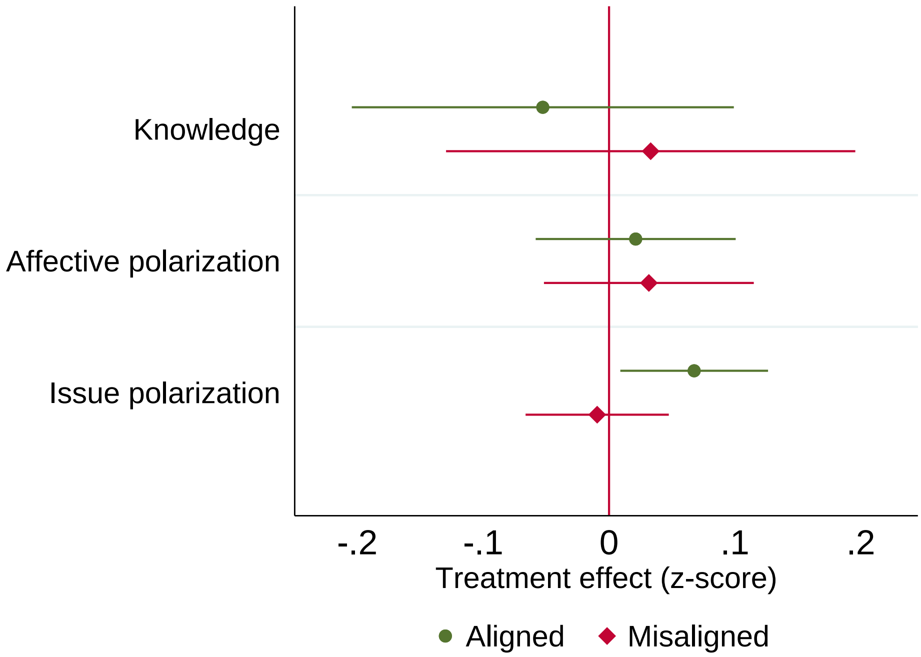
## B.11 Endline Results

Figure B.19: Effects on Perceptions and News Satisfaction at Endline




*Note:* This figure reports the same perception and news-satisfaction outcomes as Figure 2.5, measured in the endline survey. All outcomes are standardized. Error bars represent 95% confidence intervals.

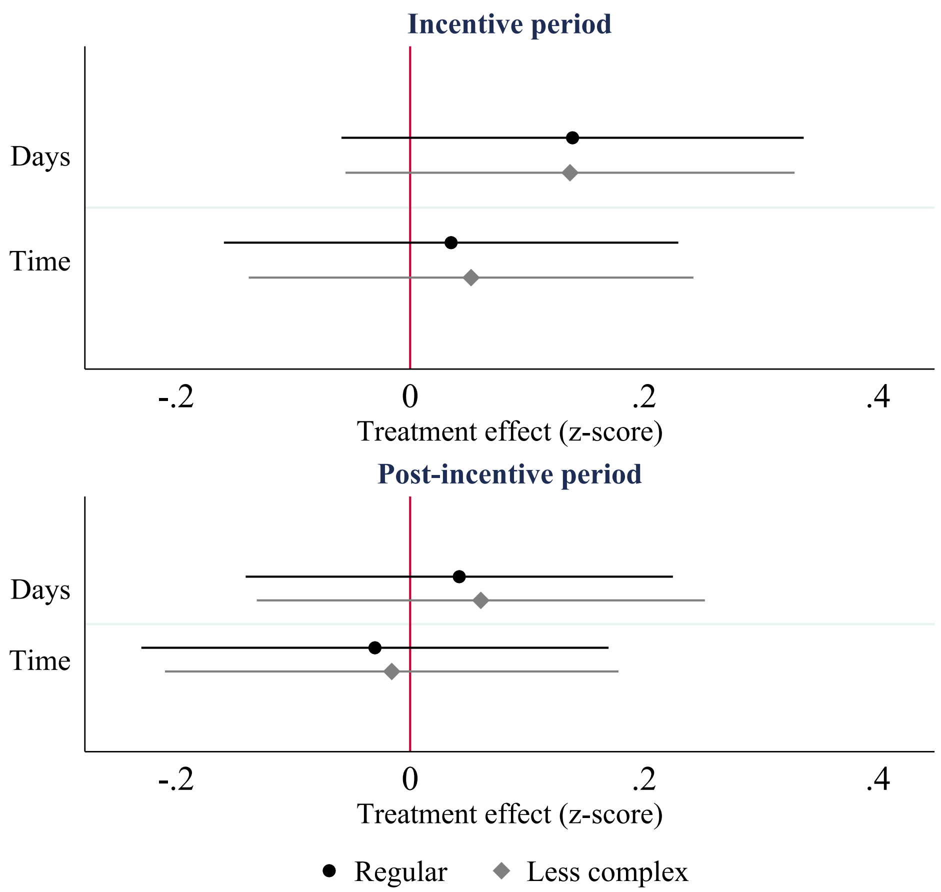
Figure B.20: Effects on Knowledge and Polarization at Endline




*Note:* This figure reports the same knowledge and polarization outcomes as Figure 2.7, measured in the endline survey. All outcomes are standardized. Error bars represent 95% confidence intervals.

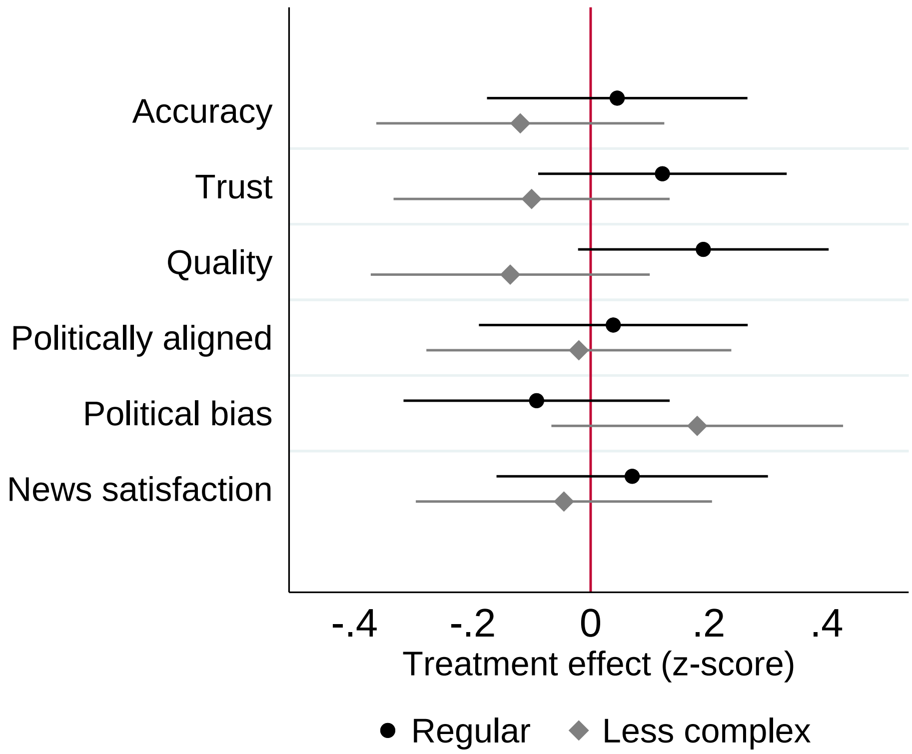
### B.12 Comparison of Control Conditions

Figure B.21: Differences in News Consumption across neutral control conditions



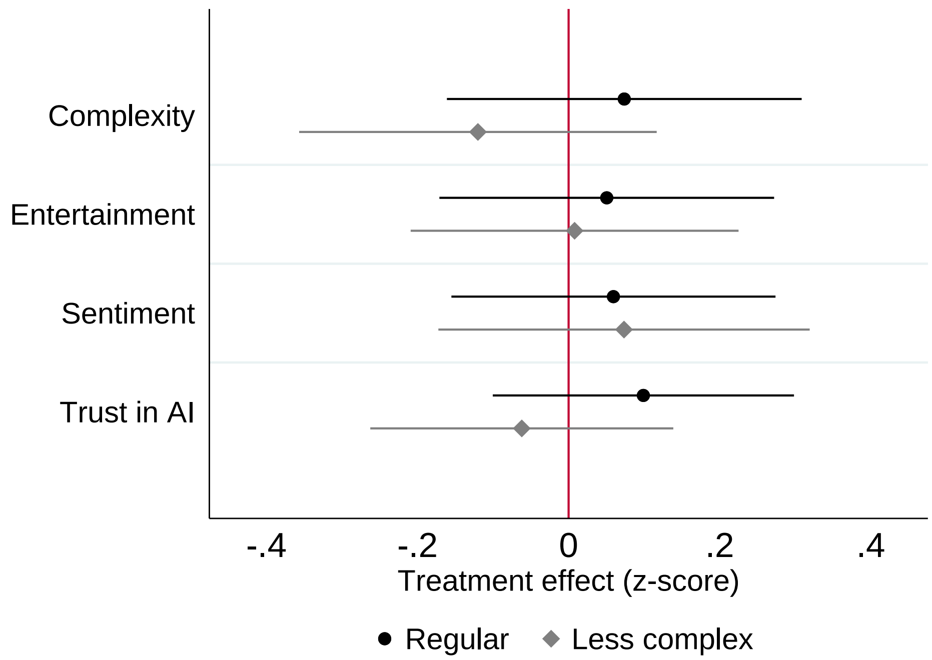

*Note:* This figure shows the effect of regular and less complex versions compared to facts-only on news consumption in the incentive period (top panel) and post-incentive period (bottom panel). News consumption is measured as the share of days using the app and winsorized daily time spent using the app. Both outcomes are standardized. Error bars represent 95% confidence intervals.

Figure B.22: Differences in Perceptions and News Satisfaction across neutral control conditions




*Note:* This figure compares the regular and less complex control conditions to facts-only for the same perception and news-satisfaction outcomes shown in Figure 2.5. All outcomes are standardized. Error bars represent 95% confidence intervals.

Figure B.23: Differences in Perceptions of Complexity, Entertainment, and Sentiment across neutral control conditions

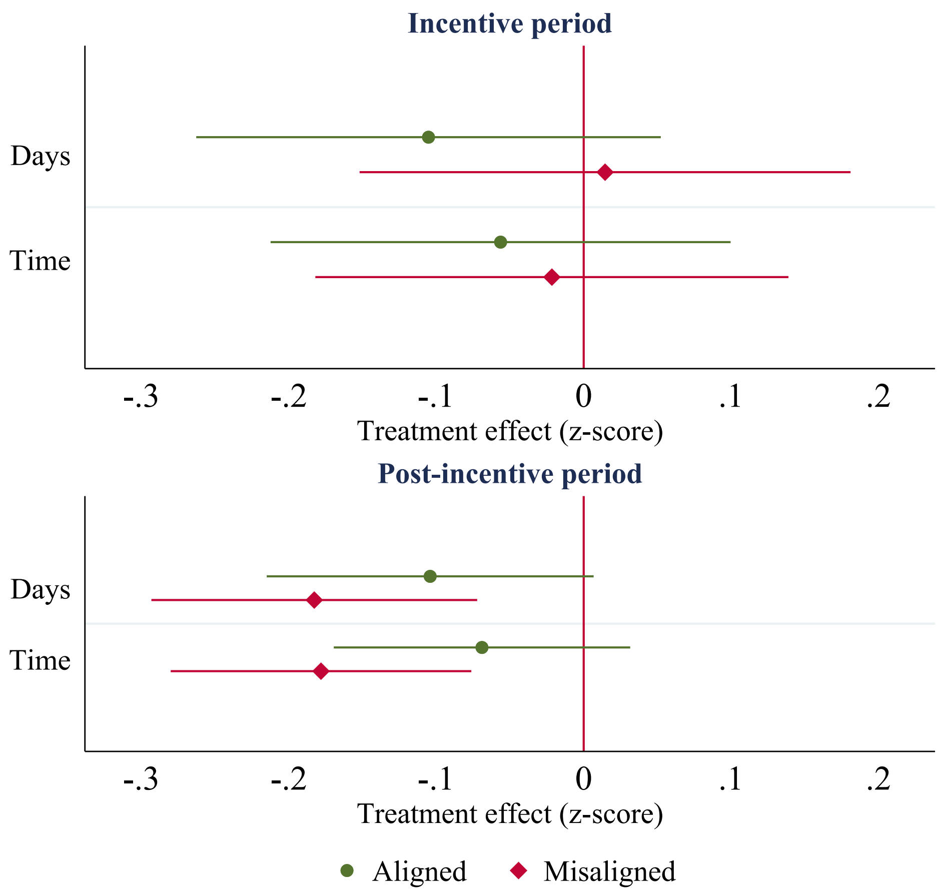



*Note:* This figure compares the regular and less complex control conditions to facts-only for the same secondary perception outcomes shown in Figure B.18. All outcomes are standardized. Error bars represent 95% confidence intervals.

## B.13 Different Political Alignment Definition

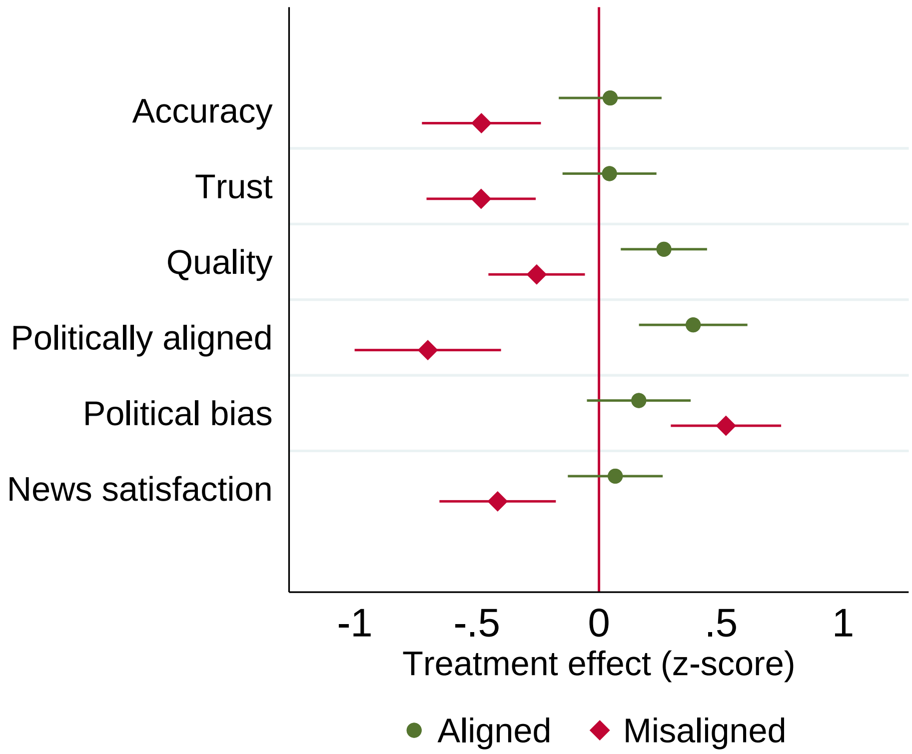
Below we describe our main results for a different definition of politically aligned news. In particular, we define political alignment based on people's response to the following question: "Are you conservative or liberal?". The response options are the following: Very conservative, conservative, neither conservative nor liberal, liberal, and very liberal. Respondents who say that they are neither conservative nor liberal are excluded from this analysis. The exhibits below confirm the patterns we uncover in our main specifications.

Figure B.24: Effects on News Consumption (Alternative Alignment Definition)



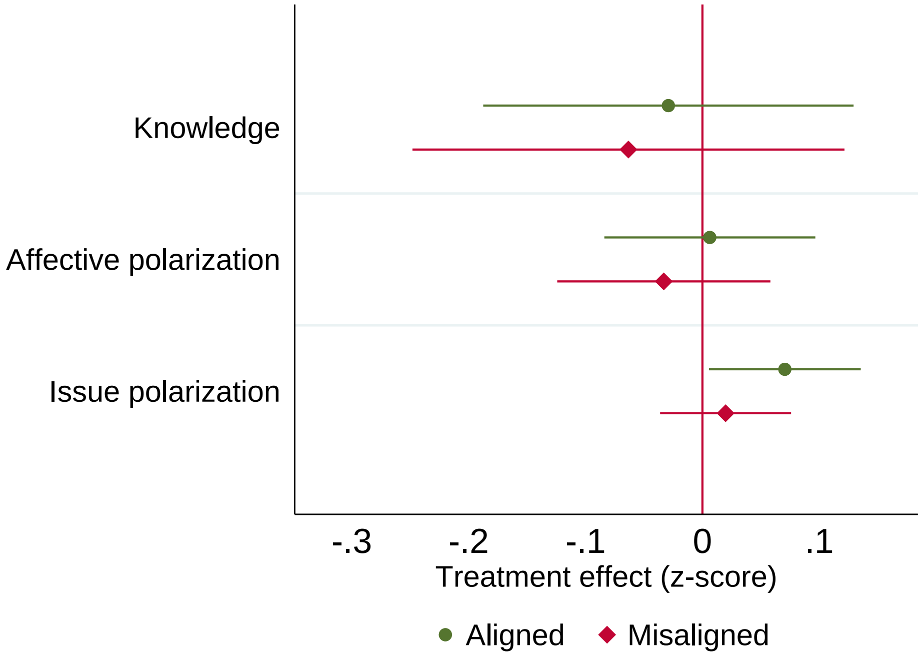

*Note:* This figure shows the effect of aligned and misaligned versions on news consumption in the incentive period (top panel) and post-incentive period (bottom panel). News consumption is measured as the share of days using the app and winsorized daily time spent using the app. Both outcomes are standardized. Error bars represent 95% confidence intervals.

Figure B.25: Effects on Perceptions and News Satisfaction




*Note:* This figure shows the effect of aligned and misaligned versions on perceptions of the app and news satisfaction. *Accuracy*, *Trust*, and *Quality* measure respondent ratings of the articles in the app. Political ratings are based on respondent ratings of articles on a 5-point scale from "very right-wing biased" to "very left-wing biased". Under this alternative specification, *politically aligned* is defined using respondents' answers to the question "Are you conservative or liberal?" and excludes respondents who answer "neither conservative nor liberal." *Political bias* is an indicator for rating it as biased in any direction (as compared to "not biased"). *News satisfaction* is respondents' reported satisfaction with the news they have been reading over the past week. All outcomes are standardized. Error bars represent 95% confidence intervals.

Figure B.26: Effects on Knowledge and Polarization




*Note:* This figure shows the effect of aligned and misaligned versions on knowledge and polarization. *Knowledge* refers to the score in a quiz of 15 questions based on recent news stories. *Affective polarization* is an index combining feeling warmly toward Republicans and Democrats and hypothetical division of $100 between a Democrat and Republican voter. *Issue polarization* is an index combining policy views on immigration, firearms, international trade, racial discrimination by police, and sexual harassment. All outcomes are standardized. Error bars represent 95% confidence intervals.

## B.14 User Experience Data

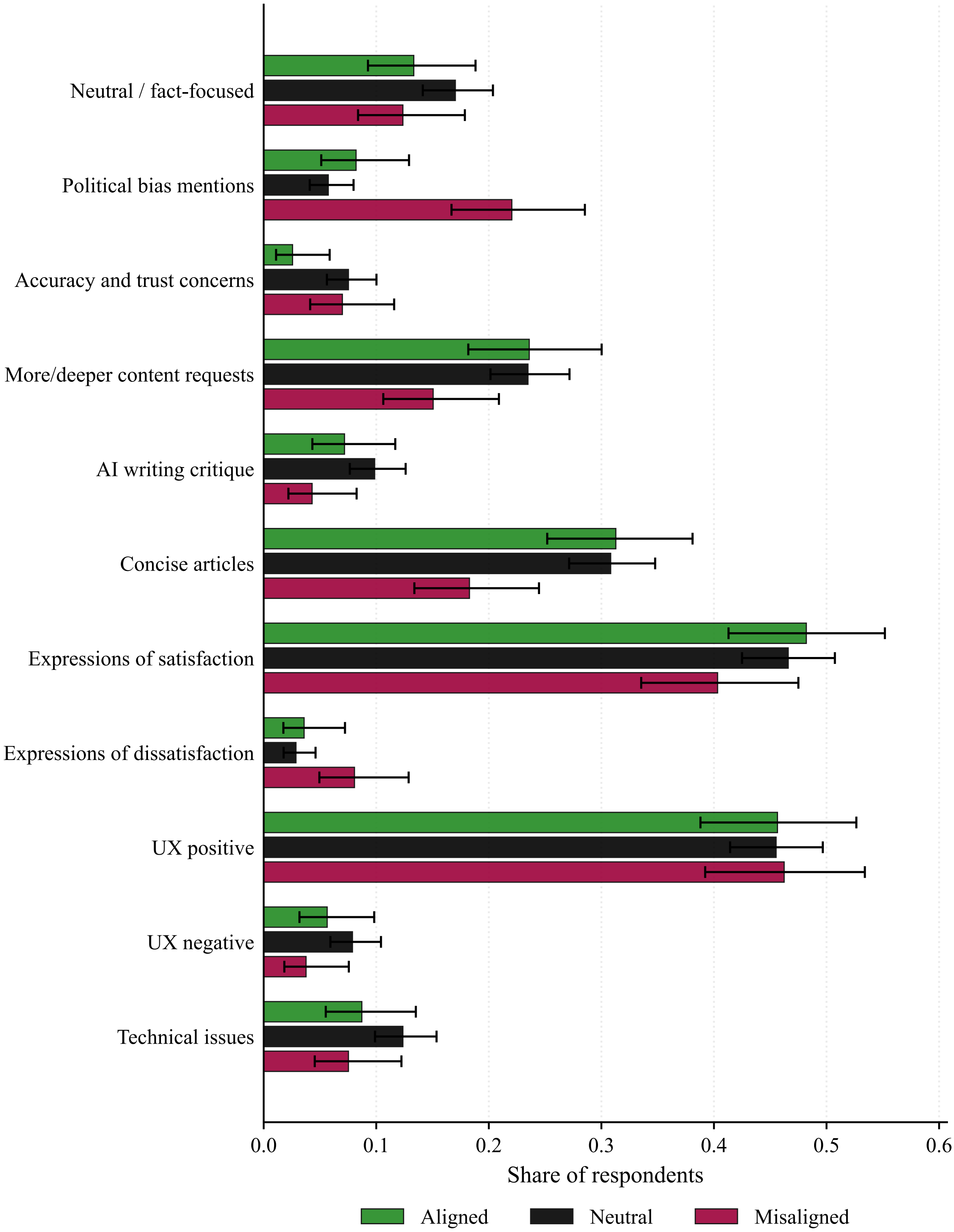
To better understand how respondents experienced the app, we analyze the open-ended question administered in the midline survey, in which respondents described their experience using the app in their own words. We first used OpenAI's `o3` model to conduct an inductive qualitative analysis of the responses and generate an initial set of themes, code definitions, and illustrative examples. We then manually reviewed and refined this output to construct the final coding scheme reported in Table B.14. Using this finalized scheme, we applied the same model to code the full set of responses in batches. For each response, the model returned structured output containing the assigned codes, a short justification, and optional notes on ambiguities or edge cases. Responses could receive multiple codes.

Table B.14: Open-Ended Response Coding Scheme

| Category | Explanation | Example |
|---|---|---|
| Neutral / fact-focused | Praise for factual, unbiased reporting style | "I found the articles fact focused with no bias or inflammatory language." |
| Political bias mentions | Claims of political bias, ideological slant, or partisan content | "I had issues with the right-leaning articles that continued to be spewed..." |
| Accuracy and trust concerns | Doubts about reliability, need to fact-check, misinformation worries | "Made the articles feel more fake or less accurate at times." |
| More/deeper content requests | Requests for more detailed articles, additional topics, broader coverage | "Only suggestion... add more articles from each category on the home page." |
| AI writing critique | Criticism of artificial, robotic, or poor writing quality | "Every article sounded like it was written by AI and that was a turnoff." |
| Concise articles | Appreciation for short, quick-read format and brevity | "Most of the articles were within 1-3 minute reads so it didn't take long to get caught up." |
| Expressions of satisfaction | Explicit positive overall assessment and intention to keep using | "I liked the app and would like to keep it installed on my device." |
| Expressions of dissatisfaction | Negative overall assessment or intention to stop using | "Will be deleting the app." |
| UX positive | Praise for ease of use, simple navigation, clean design, fast loading | "The app was simple and easy to follow and there were no bloated ads." |
| UX negative | Requests for better visuals, more appealing design, additional UI elements | "I did not like the design and the latency... black and white look... not appealing." |
| Technical issues | Any functional problems, crashes, errors, or technical malfunctions including notification issues | "When I clicked a push notification... unable to go to the homepage after reading the article." |

*Note:* The table reports the coding scheme used to classify respondents' open-ended comments about the app. Responses could receive multiple codes. The table lists the substantive codes used in the appendix analyses and provides a short definition and an illustrative example for each code.

Figure B.27: Analysis of Open-Ended Responses




*Note:* This figure shows the share of respondents in each condition whose open-ended response was coded as containing the indicated theme. Responses can receive multiple codes, so percentages need not sum to 100 across themes. Error bars represent 95% confidence intervals.

# Appendix C: Measuring the Structure of Household Macroeconomic Expectations

## C.1 Survey Instruments and Variable Construction

Tables C.3 to C.5 summarize the source survey items used in the paper. The tables record the source item, the paper measure, the response scale, the transformation used in the analysis, and the role of the variable. They are condensed summaries of the public survey documentation rather than reproductions of the full questionnaires. The notes in this section state the key mapping and construction decisions before the detailed tables.

The construction details are part of the measurement evidence. I study correlations among reported expectations, so a sign convention or scale choice can change the meaning of the measure even when the raw survey item is familiar. The following notes record the relevant transformation before any validation or behavioral interpretation. In particular, probability answers, density-implied inflation means, signed open-ended point forecasts, and ECB quantitative expectations are not interchangeable measurement formats.

#### Instrument mapping

All SCE measurement inputs use 12-month expectations. The three probability inputs share the same percent-chance format: respondents report the chance that unemployment, saving-account interest rates, or stock-market prices will be higher 12 months from now. The inflation benchmark is different by design. It uses the public density-implied mean from the one-year inflation-bin question, because that field uses the full reported probability distribution and avoids using the open-ended point forecast as the benchmark input. The signed inflation point forecast remains a robustness input and a check on the sign convention. The same inflation-bin question also supports a probability-format robustness input: the public files report the ten underlying bins, so the probability mass assigned to inflation above a fixed threshold puts inflation on the same 0–100 percent-chance scale as the three probability inputs. The threshold version asks about inflation exceeding a fixed value rather than inflation being higher than it is now, so it is the closest available analogue to the three higher-than-now probability questions, not an identical instrument.

The SCE point-forecast fields used as robustness or outcome variables are already signed in the public files. The signed inflation point forecast is not multiplied again by the direction variable. The same rule applies to spending: the public signed expected spending-growth field is used directly. This matters because multiplying an already signed public field by the direction question would mechanically create wrong signs for a subset of observations.

The SCE spending outcome is also a 12-month expectation. The tail treatment in the spending section is an outcome rule that leaves the measurement inputs untouched: the primary outcome drops respondent-month observations with absolute expected spending growth above 50 percentage points, and the robustness outcome winsorizes within survey month. This rule is kept outside the measurement construction because the comovement index is built only from the four benchmark macro expectations.

The additional SCE prediction checks use four secondary outcomes. The signed household-income growth field is used directly and trimmed at absolute values above 50 percentage points, paralleling the spending tail rule. Expected future credit access is recoded so that higher values mean harder expected credit. Debt-payment risk and job-loss risk are probability answers and are kept on the 0–100 scale after setting values outside that range to missing. Job-loss expectations are observed only for respondents routed into that question, so the corresponding check is an eligible-employed-respondent association rather than an all-household outcome.

The ECB CES mapping supports a cross-survey comparison built from imperfect analogues of the SCE instrument. The headline ECB inflation input is the open-ended expected percentage change in prices in general. The appendix also reports an inflation-elicitation robustness check that replaces this open-ended input with a midpoint-implied mean from the CES probabilistic inflation bins, using only complete bin vectors in the 0–100 range that sum to 100. The unemployment input is the expected unemployment-rate level minus the perceived current unemployment-rate level, because the closest SCE measure is a directional expectation that unemployment will be higher. The interest-rate input is expected mortgage rates, using the current CES item where available and its pre-June-2022 predecessor before the replacement; this is an imperfect analogue to the SCE saving-account-rate probability. The real-activity input multiplies expected economic growth by -1 so that higher values denote weaker expected activity. Home-price expectations are documented because they are the closest monthly CES asset-price item, but they are kept secondary because they are not a stock-price analogue.

The ECB adverse macro-comovement index uses a coherent adverse sign convention rather than the SCE probability sign convention. Higher expected inflation, higher expected unemployment, higher expected mortgage rates, and weaker expected real activity all point toward a less favorable macro assessment. This makes the ECB measure interpretable within the CES, but it also limits the cross-survey claim: the CES exercise asks whether within-person macro comovement appears in another public panel, not whether the exact SCE probability index can be replicated directly in the CES.

**Construction details**

The SCE construction proceeds in a fixed order. I construct a cleaned respondent-month panel from the raw public files, checking for duplicate respondent-months before forming any measurement panel. The exact-12 benchmark then keeps respondent-months with nonmissing values for the density-implied inflation mean, the unemployment probability, the saving-rate probability, and the stock-price probability, and retains respondents with exactly 12 such complete observations. No spending variable, validation result, or behavioral coefficient enters this sample definition.

The common-support sample is formed after the exact-12 panel is fixed. Respondents with zero raw within-person variance in any of the four benchmark inputs are removed because at least one pairwise raw-Pearson correlation would otherwise be undefined. As a result, the exact-12 benchmark has 6,659 respondents while the raw-Pearson common-support sample used in the main SCE measurement tables has 6,550 respondents. The same common-support convention is used when comparing raw Pearson, survey-month standardized Pearson, and raw Spearman variants in the SCE measurement tables.

Table C.1: SCE sample construction

| Sample | Respondents | Respondent-months | First month | Last month |
|---|---|---|---|---|
| Raw SCE public microdata | 23,886 | 180,268 | – | – |
| Complete benchmark inputs | 23,518 | 175,746 | – | – |
| Exact-12 benchmark | 6,659 | 79,908 | 2013m6 | 2025m4 |
| Exact-12 common support | 6,550 | 78,600 | 2013m6 | 2025m4 |
| Consecutive exact-12 | 6,587 | 79,044 | 2013m6 | 2025m4 |
| At-least-12, first 12 | 7,096 | 85,152 | 2013m6 | 2025m4 |

*Note:* The benchmark keeps respondents with exactly 12 complete observations for the four measurement inputs. The common-support sample additionally removes respondents with zero raw within-person variance in any benchmark input.

The alternative SCE sample windows are constructed only as robustness checks. The consecutive exact-12 sample requires the 12 complete observations to be calendar-consecutive. The at-least-12, first-12 sample keeps respondents with at least 12 complete observations and uses their first 12 complete observations. These alternatives test whether the main probability-versus-inflation ranking depends on the exact-12 rule.

Table C.2 summarizes the benchmark SCE inputs used in the exact-12 panel. The mix of percentage-point and percent-chance scales is the reason the main text starts from scale-free correlations rather than raw covariances.

Table C.2: Benchmark expectation inputs in the exact-12 panel

| Input | *N* | Mean | SD | p5 | Median | p95 |
|---|---|---|---|---|---|---|
| Inflation density mean | 79,908 | 4.09 | 4.97 | -0.97 | 3.00 | 12.01 |
| Pr(unemployment higher) | 79,908 | 37.68 | 23.45 | 5.00 | 39.00 | 80.00 |
| Pr(saving rates higher) | 79,908 | 33.05 | 26.15 | 0.00 | 30.00 | 80.00 |
| Pr(stock prices higher) | 79,908 | 42.27 | 23.20 | 5.00 | 49.00 | 80.00 |

*Note:* Inflation is the density-implied one-year mean. The other three inputs are subjective probabilities on the 0–100 percent-chance scale that the relevant outcome will be higher 12 months ahead.

The ECB CES construction follows the same logic where the survey permits it. Monthly CES files are harmonized into a respondent-round file, hidden nonresponse codes are set to missing, and candidate inputs are defined before the comparison exercise. The primary ECB panel keeps respondents with at least 12 complete observations for expected inflation, expected unemployment-rate change, expected mortgage rates, and expected economy weakness, then uses the first 12 complete observations. The resulting first-12 panel has 41,200 respondents before finite-correlation restrictions. The raw-Pearson comparison row has 38,885 finite respondents; country-wave standardized, country-wave winsorized, and raw Spearman rows are reported as additional variants. Because country-wave standardization makes within-person correlations finite even for respondents with no raw within-person movement, the comparison table also reports a common-support row that recomputes the standardized variant on the raw-Pearson finite sample. The inflation-elicitation robustness panel applies the same first-12 rule after replacing the inflation requirement with valid probabilistic-bin inflation means.

Table C.3: SCE survey instruments and measures: benchmark and spending

| Source item | Paper measure | Instrument content and scale | Construction | Role |
|---|---|---|---|---|
| Q9 / Q9_ mean | Density-implied inflation expectation | One-year inflation density question: respondents allocate probabilities across inflation and deflation bins; the public file reports a density-implied mean. Scale: Ten probability bins that sum to 100; constructed density mean in annual percentage points. | Use the public density-implied mean as the benchmark inflation input. | Benchmark measurement input |
| Q4new | Unemployment-higher probability | Percent chance that the U.S. unemployment rate will be higher 12 months from now. Scale: 0-100 probability response using ruler and entry box. | Set values outside 0-100 to missing. | Benchmark measurement input |
| Q5new | Saving-rate-higher probability | Percent chance that the average saving-account interest rate will be higher 12 months from now. Scale: 0-100 probability response using ruler and entry box. | Set values outside 0-100 to missing. | Benchmark measurement input |
| Q6new | Stock-price-higher probability | Percent chance that average U.S. stock-market prices will be higher 12 months from now. Scale: 0-100 probability response using ruler and entry box. | Set values outside 0-100 to missing. | Benchmark measurement input |
| Q26v2 / Q26v2part2 | Expected spending growth | Expected direction and percent size of total household-spending change over the next 12 months. Scale: Direction question plus nonnegative percent magnitude; the public field is already signed. | Use the public signed field; trim absolute values above 50 percentage points in the primary outcome and use month-level 1/99 winsorization as robustness. | Behavioral outcome |
| Q8v2 / Q8v2part2 | Signed inflation point forecast | Expected direction and rate of inflation or deflation over the next 12 months. Scale: Direction question plus nonnegative percent magnitude; the public field is already signed. | Use only as a robustness and coding-check input; do not multiply again by the direction field. | Robustness input |

*Note:* Source document is the public NY Fed SCE core questionnaire. Question content is summarized for compactness. The relevant public SCE point-forecast fields are used as signed in the microdata.

Table C.4: SCE survey instruments and measures: secondary and diagnostic variables

| Source item | Paper measure | Instrument content and scale | Construction | Role |
|---|---|---|---|---|
| Q25v2 / Q25v2part2 | Expected household-income growth | Expected direction and percent size of total household-income change over the next 12 months. Scale: Direction question plus percent magnitude; the public field is already signed. | Use the public signed field directly; trim absolute values above 50 percentage points in the personal-risk prediction check. | Secondary prediction outcome |
| Q29 | Expected credit difficulty | Expected future difficulty of obtaining credit or loans 12 months from now. Scale: Five-point scale from much harder to much easier. | Keep valid values from 1 to 5 and reverse the scale so higher values mean harder expected credit. | Secondary prediction outcome |
| Q30new | Debt-payment risk | Percent chance of being unable to make a debt payment over the next 3 months. Scale: 0-100 probability response using ruler and entry box. | Set values outside 0-100 to missing. | Secondary prediction outcome |
| Q13new | Job-loss risk | Percent chance of losing the current or main job over the next 12 months. Scale: 0-100 probability response for eligible employed respondents. | Set values outside 0-100 to missing; non-eligible respondents are missing by survey routing. | Secondary prediction outcome |
| Q3 | Moving-residence probability | Percent chance of moving to a different primary residence over the next 12 months. Scale: 0-100 probability response using ruler and entry box. | Set values outside 0-100 to missing; use only in the auxiliary probability placebo diagnostic. | Auxiliary probability diagnostic |
| Q14new | Voluntary job-quit probability | Percent chance of voluntarily leaving the current or main job over the next 12 months. Scale: 0-100 probability response for eligible employed respondents. | Set values outside 0-100 to missing; non-eligible respondents are missing by survey routing. | Auxiliary probability diagnostic |
| Q22new | Job-finding probability after job loss | Percent chance of finding an acceptable job within 3 months if the respondent lost the current or main job this month. Scale: 0-100 probability response for eligible employed respondents. | Set values outside 0-100 to missing; non-eligible respondents are missing by survey routing. | Auxiliary probability diagnostic |
| Q9 / Q9_iqr | Inflation-density interquartile range | Interquartile range of the density implied by the one-year inflation-bin question. Scale: Density IQR in annual percentage points. | Set negative values to missing; used only in the uncertainty-mechanism diagnostic and not in the benchmark sample. | Mechanism diagnostic input |
| Q9 / Q9_ bin1-Q9_ bin10 | Inflation density mass above 4 or 2 percent | Density probability mass on inflation above 4 percent (bins 1-3) or above 2 percent (bins 1-4). Scale: Sum of 0-100 probability bins. | Valid only if all ten bins are in 0-100 and sum to 100 within rounding tolerance; inflation-elicitation robustness only. | Robustness input |

*Note:* Source document is the public NY Fed SCE core questionnaire. Question content is summarized for compactness. The relevant public SCE point-forecast fields are used as signed in the microdata.

Table C.5: ECB CES survey instruments and measures

| Source item | Paper measure | Instrument content and scale | Construction | Role |
|---|---|---|---|---|
| C1120 | Expected inflation | Expected percent change in prices in general over the next 12 months. Scale: Numeric percent response from -100 to 100 with one decimal place. | Use the cleaned expected-inflation response; robustness variants include country-wave standardization and winsorization. | ECB comparison input |
| C1150_* / C1152_* | Inflation-bin midpoint mean | Expected distribution of price changes in general over the next 12 months. Scale: Probabilistic bins that allocate 100 points across price-increase and price-decrease intervals; the source items use 8-bin and 10-bin versions. | Use only as an inflation-elicitation robustness input; require complete bins in 0-100 that sum to 100 and convert to a midpoint-implied mean. | ECB robustness input |
| C4030 / C4031 | Expected unemployment-rate change | Current unemployment rate and expected unemployment rate 12 months ahead in the respondent's country. Scale: Numeric percent response from 0 to 100 for each unemployment-rate level. | Construct expected unemployment-rate change as the expected level minus the perceived current level. | ECB comparison input |
| C5113 / C5111 | Expected mortgage rate | Expected interest rate on mortgages 12 months ahead in the respondent's country. Scale: Numeric percent response with one decimal place; the source items have different upper ranges. | Use the current item where available and the predecessor item as fallback; the current item replaced its predecessor in June 2022. | Imperfect SCE saving-rate analogue |
| C4020 | Expected economic weakness | Expected percent growth or shrinkage of the economy over the next 12 months. Scale: Numeric percent response from -100 to 100 with one decimal place. | Multiply expected economic growth by -1 so higher values mean weaker expected real activity. | ECB comparison input |
| C2120 | Expected home-price growth | Expected percent change in the price of the respondent's current home over the next 12 months. Scale: Numeric percent response from -100 to 100 with one decimal place. | Keep as a secondary asset-price candidate; it is not a direct stock-price analogue. | Secondary asset-price candidate |

*Note:* Source documents are the public ECB CES Microdata Guide and Methodological Guide. Question content is summarized for compactness. The ECB variables are related inputs for the cross-survey comparison, not one-for-one analogues to the SCE probability index.

## C.2 Illustrative Mechanisms for Joint Expectation Movements

Section 3.3 uses this table to motivate the sign-agnostic measurement design. The table gives economic and measurement reasons that the same four survey items can move together in different ways. It does not assign respondents to the listed mechanisms.

Table C.6: Illustrative mechanisms behind joint expectation movements

| Interpretation | Possible comovement implication | Sources of sign ambiguity |
|---|---|---|
| Demand expansion | Inflation, rate, and stock-price expectations rise while unemployment risk falls if news is read as stronger nominal and real activity. | The same news can instead be read as price pressure or future tightening, so no uniform sign pattern follows. |
| Cost-push pressure | Inflation rises with unemployment risk and rate expectations if inflation is read as adverse supply or purchasing-power pressure. | The same inflation news can be read as temporary, policy-driven, or unrelated to local labor risk. |
| Inflation-driven tightening | Inflation and rate expectations rise together if policy is expected to respond, while unemployment risk rises and stock-price expectations fall. | Households report saving-account rates, not policy rates, and differ in how much policy reaction they perceive. |
| Financial stress | Unemployment risk and rate expectations comove with weaker asset-market expectations if respondents perceive tighter credit or risk premia. | Expected policy easing would reverse the rate sign; stock-price beliefs can also reflect expected nominal growth or recovery. |
| Broad macro uncertainty or perceived change | Several event probabilities rise together when respondents perceive that many macro outcomes are likely to move. | Generates positive probability comovement without being an optimism-pessimism factor. |
| Probability-response behavior | Common use of focal or high probabilities creates comovement across probability answers. | A competing measurement interpretation, tested directly through focal responses, levels, timing, and reducibility. |

*Note:* The table gives economic and measurement reasons that the same four survey items can move together in different ways. It does not assign respondents to the listed mechanisms.

## C.3 Reduced-Form Algebra Behind the Slope Prediction

Section 3.3 states the one-variable-slope decomposition, Equation (3.1); this appendix states the assumptions behind it and the limits on its interpretation. The decomposition is not a structural model of household reasoning; it shows why a one-variable expectation coefficient can depend on how that expectation comoves with the rest of the respondent's expectation vector.

The vector $z_{it}$ standardizes respondent $i$'s repeated macro expectations within respondent, so each component has mean zero and variance one over the retained panel window, and $\tilde{y}_{it}$ is another reported expectation or outcome demeaned within respondent. The coefficients $\lambda_i$ in the approximation $\tilde{y}_{it} = \lambda_i' z_{it} + u_{it}$ are not interpreted as structural preference or belief parameters; they only summarize the possibility that several reported expectations enter a respondent's reduced-form mapping to another survey answer. Because $z_{ijt}$ has unit within-respondent variance, the population slope from the one-variable regression equals the covariance on the left side of Equation (3.1). Even in this stripped-down representation, the reduced-form association between one expectation and another reported variable depends on the respondent's correlation structure among expectations.

The implication for the paper is narrow but useful. If two households have the same average inflation expectation and the same marginal variance of inflation expectations, the same one-variable inflation coefficient need not summarize the same belief system when inflation comoves differently with unemployment, interest-rate, or asset-market expectations. Correlation patterns capture information that levels alone do not contain. I measure those correlations directly and validate which summaries are stable enough to use.

The same algebra also limits the claim. The correlations $\rho_{i,jk}$ do not reveal which expectation causes another expectation to move, which news respondents process, or whether a respondent endorses a particular macroeconomic model. They summarize repeated alignment in survey answers. The empirical sections therefore use language such as comovement, reduced-form mapping, predictive content, and cross-survey comparison, and avoid the language of causal effects or structural transmission.

## C.4 Measurement Formula Details

This section defines the formulas behind the validation diagnostics. The primitive within-household correlations $\rho_{i,jk}$ and the reduced signed-comovement indices are defined in Section 3.4, Equation (3.2). Correlations are computed only for respondents with positive within-person variance in all four benchmark inputs over the retained 12 waves; respondents with a zero-variance input can still be informative about levels, but they are excluded from the correlation tables, and this condition coincides with the common-support sample that requires all six benchmark correlations to be finite.

The SCE variants keep the same target but alter how the primitive correlations are computed. The survey-month standardized Pearson variant first standardizes each input within survey month and then computes the within-respondent correlation, removing common month-level location and scale differences. The Spearman variant replaces each respondent's 12 answers for a given input with within-respondent ranks before applying the same formula, which reduces sensitivity to tail values.

#### Null and reliability diagnostics

The time-permutation null asks whether the observed average comovement could arise from the marginal time series alone. For each respondent and each benchmark input, I randomly permute the 12 retained observations over time, preserving that respondent's marginal distribution for the input while breaking within-month alignment across inputs. Primitive correlations and reduced indices are recomputed in each permuted sample. The null tables compare the observed average index to the distribution of average indices across permutation iterations. This is a noise benchmark for temporal alignment.

Split-half reliability asks whether respondent rankings survive when the same 12-wave histories are divided into shorter histories. The first-last diagnostic computes the measure separately in waves 1–6 and waves 7–12. The odd-even diagnostic computes it in alternating waves. The random-split diagnostic averages reliability over many respondent-specific random 6/6 splits. For a measure $M_i$, reliability is the cross-respondent correlation between the two split estimates:

$$\mathrm{Rel}(M) = \mathrm{Corr}_i(M_i^{(a)}, M_i^{(b)}).$$

The Spearman-Brown adjustment reported in the tables is

$$\mathrm{SB}(\mathrm{Rel}) = \frac{2\,\mathrm{Rel}}{1 + \mathrm{Rel}},$$

which gives the standard projection from half-length reliability to full-length reliability under classical assumptions. I report both the raw split correlation and the adjusted statistic because the raw statistic is closer to the data and the adjustment is easier to compare to familiar reliability benchmarks.

The jackknife diagnostic is different. It recomputes the full measure after leaving out one retained wave at a time and records how closely the leave-one-wave-out ranking matches the full-12 ranking. This is not a substitute for split-half validation because the samples overlap heavily. It is a mechanical check that the respondent ordering is not driven by a single month in the retained panel.

**Reducibility checks**

The reducibility exercises project each reduced measure on observed respondent characteristics and survey-behavior controls. The goal is to ask whether the proposed measure is mostly a relabeling of average expectations, focal response behavior, timing, or demographics, not to estimate structural determinants of the index. The main blocks are expectation levels, response-style variables, timing and panel-gap variables, demographics, education, income, and numeracy.

The reported $R^2$ values have a measurement interpretation. A high value would mean the index adds little beyond existing observed variables. A value near zero would mean the index is empirically distinct from the observed controls, though not necessarily reliable. The probability signed-comovement index falls between these extremes: observed controls explain a meaningful share, especially expectation levels and demographics, but a large residual component remains. The observed $R^2$ should also be read against the index's split-half reliability: under classical measurement error the controls capture roughly half of the index's reliable variance, and the unexplained share of observed variance mixes unmeasured stable comovement with estimation noise. The index is informative, but it is not a pure latent trait.

Focal answers at 0, 50, and 100 can create artificial similarity across probability questions, so they are measured directly and included in the reducibility checks. Response style matters but does not absorb the probability-comovement measure.

## C.5 Validation Evidence and Interpretation Boundaries

The evidence is cumulative: construction checks, noise benchmarks, reliability checks, reducibility checks, cross-survey comparisons, and prediction exercises jointly constrain the interpretation. Table C.7 summarizes those constraints.

Table C.7: Validation hierarchy and interpretation boundaries

| Evidence layer | Supports | Does not support |
|---|---|---|
| Benchmark construction | The exact-12 SCE design defines the measure before behavioral analysis and gives a common panel length for split-half validation. | The exact-12 sample is not a population reweighting scheme, and it does not imply that longer histories or alternative panels are irrelevant. |
| Noise benchmarks | Time-permutation nulls show that observed probability comovement is not just a mechanical consequence of each respondent's marginal answer series. | The null is not a model of respondent behavior, information arrival, or rational updating. |
| Reliability checks | Split-half, random-split, and jackknife diagnostics justify using the continuous probability signed-comovement index as a noisy respondent ranking. | The evidence is not strong enough for precise psychological traits or durable discrete type labels from 12 public survey waves. |
| Reducibility checks | Levels, timing, demographics, numeracy, and focal-response controls do not exhaust the primary measure. | The index is not a pure latent trait orthogonal to observables or survey-answering behavior. |
| ECB CES comparison | Within-person comovement of adverse macro expectations appears in another public expectations panel and survives basic country, timing, scale, and tail checks. | The CES exercise is not an exact replication of the SCE probability index and does not establish a pooled cross-survey household type system. |
| Prediction exercises | Later expectation-dynamics, spending, and personal-risk specifications map where the validated measure has predictive content and where it does not. | These rows are not causal effects and do not settle behavioral relevance in richer designs with better outcomes, experiments, or linked administrative data. |

*Note:* The table describes the interpretation of evidence reported in the main text and appendix. It does not add a separate empirical design.

The hierarchy points to a clear conclusion: the SCE probability signed-comovement index is the most defensible measure in the public SCE data. It is based on economically interpretable primitive links, clearly separated from average expectation levels, well above respondent-level time-permutation benchmarks, and more stable than discrete class assignments.

## C.6 Additional SCE Measurement Details

Table C.8 documents two response-pattern features that matter for interpreting the measurement sample. Focal shares show how often the probability inputs equal 0, 50, or 100 in the exact-12 panel. Zero-variance shares show how many respondents have no within-person movement in a benchmark input and cannot contribute all raw-Pearson correlations.

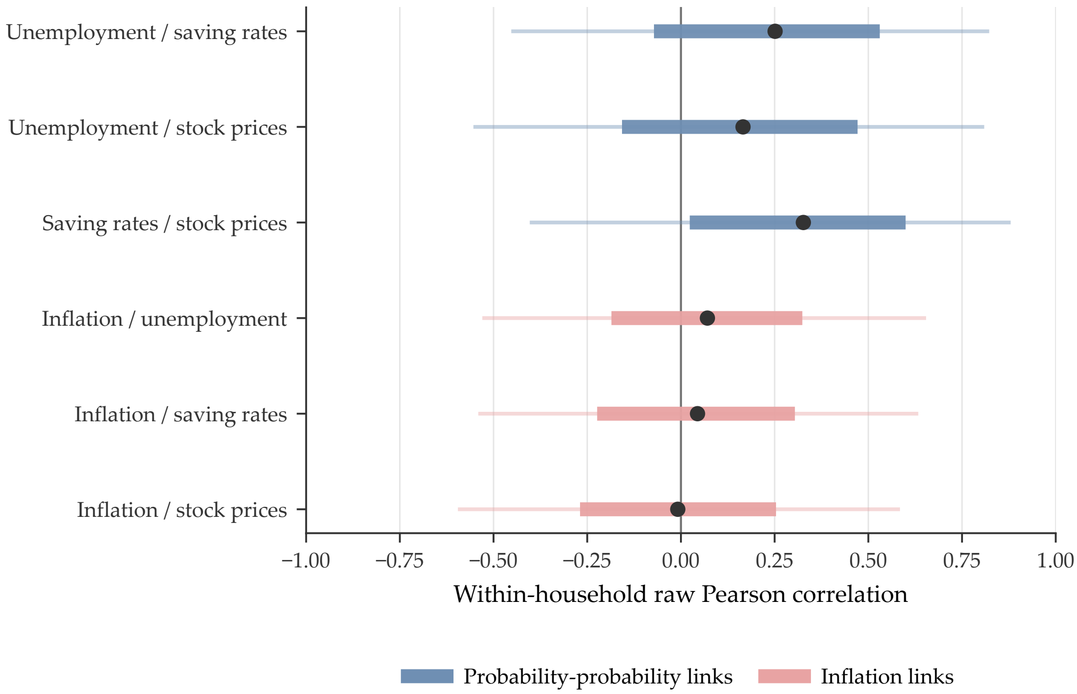
Table C.8: Focal-response and finite-correlation diagnostics

| Input | Share 0 | Share 50 | Share 100 | Zero-variance share |
|---|---|---|---|---|
| Inflation density mean | – | – | – | 0.83 |
| Pr(unemployment higher) | 1.4 | 16.8 | 1.1 | 0.33 |
| Pr(saving rates higher) | 5.6 | 11.0 | 0.8 | 0.42 |
| Pr(stock prices higher) | 1.1 | 18.1 | 0.6 | 0.29 |
| Any benchmark input | – | – | – | 1.64 |

*Note:* All entries are in percent. Focal-response shares are respondent-month shares in the exact-12 benchmark panel for probability inputs. Zero-variance shares are respondent-level shares in the exact-12 benchmark; respondents with zero raw within-person variance in any benchmark input are excluded from the raw-Pearson common-support sample.

Figure C.1 shows the distribution of the six primitive within-household correlations reported in the main text. These correlations remain the measurement foundation even though the main text emphasizes the reduced probability signed-comovement index.

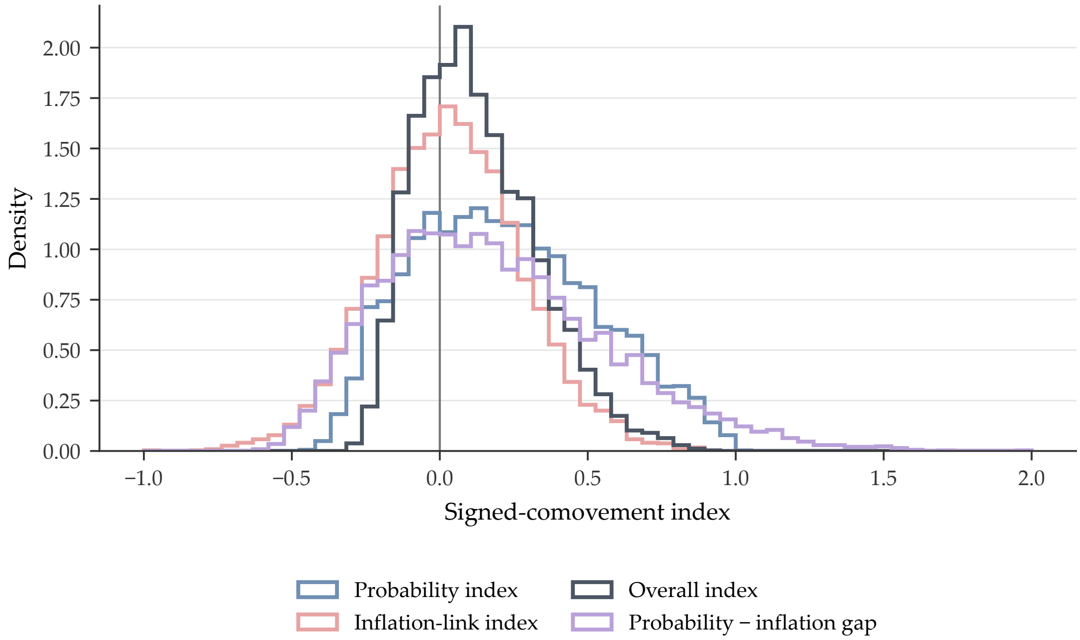
Figure C.1: Primitive SCE within-household correlations




*Note.* The figure shows raw Pearson correlations computed within exact-12 respondents in the common-support sample. Dots mark medians, thick segments mark interquartile ranges, and thin segments mark fifth-to-ninety-fifth percentile ranges.

Figure C.2 shows the reduced signed-comovement indices used to rank the SCE measures. The distributions are continuous and overlapping. This is the visual counterpart to the validation conclusion in the main text: the public SCE supports a continuous ranking more naturally than a small number of discrete household labels.

Figure C.2: SCE signed-comovement index distributions




*Note.* The figure shows raw-Pearson signed-comovement indices in the exact-12 common-support sample. Split-half reliability and reducibility checks determine which reduced measure is used as the main measure.

Table C.9 reports PCA loadings for the six standardized raw-Pearson correlations. PCA serves as a diagnostic reduction of the same correlations.

Table C.10 combines the sample-window and transformation variants for the reduced signed-comovement measures. The exact-12 raw-Pearson row is still the benchmark, but the table reports the broader robustness pattern in one place: the probability index remains well above the inflation-link index across the simple alternatives, including the first-difference row, which computes the same correlations on within-respondent changes between adjacent retained waves and is insensitive to slow drift or linear respondent-specific trends; adjacent retained waves are calendar-adjacent for the 99 percent of benchmark respondents with consecutive panels.

Table C.9: PCA loadings for the six within-household correlations

| Correlation pair | PC1 | PC2 |
|---|---|---|
| Inflation / unemployment | 0.100 | 0.486 |
| Inflation / saving rates | 0.148 | 0.653 |
| Inflation / stock prices | 0.274 | 0.495 |
| Unemployment / saving rates | 0.541 | -0.120 |
| Unemployment / stock prices | 0.598 | -0.198 |
| Saving rates / stock prices | 0.492 | -0.198 |
| Explained variance (percent) | 30.0 | 23.1 |

*Note:* PCA is run on the six standardized raw-Pearson correlations in the exact-12 common-support sample. PC1 is oriented to have a positive average loading; PC2 is oriented to load positively on the inflation-pair features. Entries are unit-norm principal-component coefficient vectors, so squared entries sum to one within each component; they are not variance-scaled correlation loadings.

Table C.10: SCE reduced measures across sample and transformation variants

| Sample | Variant | *N* | Probability | Inflation | Overall | Gap |
|---|---|---|---|---|---|---|
| Exact-12 benchmark | Raw | 6,550 | 0.222 | 0.033 | 0.128 | 0.189 |
| Exact-12 benchmark | Month-std. | 6,550 | 0.224 | 0.028 | 0.126 | 0.196 |
| Exact-12 benchmark | Spearman | 6,550 | 0.211 | 0.031 | 0.121 | 0.180 |
| Exact-12 benchmark | First difference | 6,550 | 0.205 | 0.020 | 0.113 | 0.184 |
| Consecutive exact-12 | Raw | 6,480 | 0.222 | 0.033 | 0.127 | 0.189 |
| Consecutive exact-12 | Month-std. | 6,480 | 0.224 | 0.028 | 0.126 | 0.196 |
| Consecutive exact-12 | Spearman | 6,480 | 0.210 | 0.031 | 0.121 | 0.180 |
| At-least-12, first 12 | Raw | 6,980 | 0.220 | 0.033 | 0.127 | 0.187 |
| At-least-12, first 12 | Month-std. | 6,980 | 0.224 | 0.028 | 0.126 | 0.196 |
| At-least-12, first 12 | Spearman | 6,980 | 0.210 | 0.031 | 0.121 | 0.178 |

*Note:* Rows report means of the reduced signed-comovement measures on common-support respondents for each sample window and transformation. The table combines the nearby sample-window and transformation checks behind the main measurement hierarchy. The exact-12 raw-Pearson row remains the benchmark. The first-difference row recomputes the same correlations on within-respondent changes between adjacent retained waves, so slow drift or linear respondent-specific trends cannot drive it; retained waves are calendar-adjacent for 99 percent of benchmark respondents. The three sample windows share most respondents, so several variant means coincide at three decimals.

Table C.11 checks whether this hierarchy depends on using the density-implied inflation mean rather than the signed open-ended point forecast. The comparison keeps the density-implied mean as the benchmark and then isolates the elicitation change on the exact-12 panel where both inflation fields are observed. The point-forecast rows raise the inflation-link mean somewhat, and the split reliability of the inflation-link index improves; the stronger point-forecast links survive winsorization and Spearman correlations, so tail sensitivity alone does not explain the difference between the two elicitations. The density-implied mean remains the headline input because it is the distribution-based elicitation designated as the benchmark from the outset. The check shows that this choice is conservative: the inflation links would be modestly stronger under the alternative elicitation but still far below the probability block. The bin rows address the remaining format concern directly. The density mean is the only benchmark input not on the 0–100 percent-chance scale, so weak inflation links could in principle reflect the format difference rather than the belief structure. Replacing the density mean with the probability mass the respondent's own reported density assigns to inflation above 4 percent, or above 2 percent, puts all four inputs on the same percent-chance scale. The inflation-link means remain small, 0.040 and 0.018 against a probability-index mean of 0.223, and split reliability does not improve. The weak inflation links are therefore not an artifact of mixing a percentage-point input with percent-chance inputs.

Table C.11: SCE inflation-elicitation robustness

| Specification | *N* | Prob. mean | Infl.-link mean | Overall mean | Infl.-link first/last | Infl.-link odd/even |
|---|---|---|---|---|---|---|
| Density benchmark | 6,550 | 0.222 | 0.033 | 0.128 | 0.011 | 0.152 |
| Density, common sample | 6,436 | 0.221 | 0.034 | 0.127 | 0.012 | 0.150 |
| Point forecast | 6,415 | 0.222 | 0.047 | 0.135 | 0.053 | 0.196 |
| Point forecast, winsorized | 6,417 | 0.222 | 0.048 | 0.135 | 0.051 | 0.195 |
| Point forecast, Spearman | 6,415 | 0.211 | 0.047 | 0.129 | 0.048 | 0.188 |
| Pr(inflation > 4%) from bins | 6,350 | 0.223 | 0.040 | 0.131 | -0.001 | 0.175 |
| Pr(inflation > 2%) from bins | 6,301 | 0.223 | 0.018 | 0.121 | -0.001 | 0.152 |

*Note:* Rows report means of the reduced signed-comovement measures. The first row is the headline density-implied inflation benchmark. The next rows isolate the elicitation change by using the same exact-12 panel with both inflation fields observed before finite-correlation restrictions (6,542 respondents). The reported N is the number of respondents with finite correlations for the row-specific specification. Reliability columns report split-half correlations for the inflation-link index. In the common panel, 0.92 percent of signed point forecasts exceed 50 percentage points in absolute value and 0.01 percent exceed 100 percentage points; month-level winsorization and Spearman correlations leave the main hierarchy intact. Month-specific winsorization caps can create within-person variation for respondents with otherwise constant extreme point forecasts, which is why the winsorized row's N is slightly larger; that row is a tail-sensitivity check only. The bin rows replace the density mean with the probability mass the respondent's reported inflation density assigns to inflation above 4 (above 2) percent, so the inflation input uses the same 0–100 percent-chance scale as the three probability inputs; these rows use the exact-12 panel with complete valid bins (6,659 respondents).

Table C.12 addresses the remaining measurement-noise concern. Within-person correlations are attenuated by wave-to-wave noise in each input, so the weak inflation links could in principle reflect a noisier inflation input rather than genuinely looser links. The SCE elicits inflation twice in each wave, as a density-implied mean and as a signed point forecast, and under classical measurement error the within-person covariance between the two elicitations identifies the within-person signal variance of inflation beliefs, following the duplicate-elicitation logic of Gillen et al. (2019). The duplicate covariance is positive for 89 percent of common-sample respondents, the median implied reliability of the density mean is 0.71, and the mean duplicate correlation is 0.54. Dividing the observed inflation links by the square root of the median reliability raises the inflation-link mean from 0.034 to 0.040. Dividing instead by the mean duplicate correlation, which under classical assumptions also absorbs noise in the point forecast and therefore over-corrects, raises it to at most 0.062. Both corrected values remain well below the uncorrected probability-block mean of 0.221, and the comparison is conservative because only the inflation side is corrected while the probability links are left attenuated. Winsorizing the point forecast barely changes the moments. Positively correlated noise across the two same-wave elicitations would weaken the bound reading, but the conclusion requires only that noise of plausible size cannot close a gap of this magnitude.

Table C.12: Inflation measurement noise and disattenuated inflation links

| Quantity | *N* | Mean | Median | p5 | p95 |
|---|---|---|---|---|---|
| *Panel A: Within-person duplicate-elicitation moments* | | | | | |
| Duplicate correlation, raw point forecast | 6,370 | 0.544 | 0.648 | -0.192 | 0.954 |
| Density-mean reliability, raw point forecast | 6,436 | – | 0.707 | -0.346 | 2.542 |
| Duplicate correlation, winsorized point forecast | 6,371 | 0.546 | 0.649 | -0.192 | 0.954 |
| Density-mean reliability, winsorized point forecast | 6,436 | – | 0.704 | -0.325 | 2.430 |
| Share with positive duplicate covariance | 6,436 | 0.894 | – | – | – |
| *Panel B: Observed and noise-corrected mean inflation links* | | | | | |
| | | Observed | Reliability | | Bound |
| Inflation / unemployment | | 0.066 | 0.079 | | 0.122 |
| Inflation / saving rates | | 0.042 | 0.050 | | 0.077 |
| Inflation / stock prices | | -0.007 | -0.008 | | -0.012 |
| Inflation-link index | | 0.034 | 0.040 | | 0.062 |
| Inflation-link index (winsorized duplicate) | | 0.034 | 0.040 | | 0.062 |
| Probability index (uncorrected reference) | | 0.221 | – | | – |

*Note:* Rows use the exact-12 common sample in which both same-wave inflation elicitations, the density-implied mean and the signed point forecast, are observed alongside the three probability inputs, restricted to respondents with finite benchmark correlations. Panel A reports within-person moments of the two elicitations over the 12 waves: the duplicate correlation and the implied reliability of the density mean, the within-person covariance between the two elicitations divided by the within-person variance of the density mean. Panel B divides the observed mean inflation links by the square root of the median reliability (reliability column) and by the mean duplicate correlation (bound column); under classical measurement error the duplicate correlation equals the geometric mean of the two elicitations' reliabilities and is therefore a lower bound on the square root of the density mean's reliability, the attenuation factor relevant for these correlations, so the bound column over-corrects. The reliability ratio has heavy tails from near-zero denominators, so its mean is suppressed and the correction uses the median. Only the inflation side is corrected; the probability-index reference row is left attenuated. Positively correlated elicitation noise within wave would weaken the bound interpretation. The winsorized rows replace the point forecast with its survey-month 1/99 winsorized variant.

Table C.13 reports how much of each primitive correlation is explained by focal probability-scale use. The focal controls explain little of the primitive correlations, with somewhat larger but still limited explanatory power for probability-probability links. Focal response behavior is a relevant diagnostic but well short of a full explanation of the measure.

Table C.13: Primitive-correlation reducibility to focal response controls

| Primitive correlation | $N$ | Focal $R^2$ | Residual SD ratio |
|---|---|---|---|
| Inflation / unemployment | 6,550 | 0.002 | 0.999 |
| Inflation / saving rates | 6,550 | 0.001 | 0.999 |
| Inflation / stock prices | 6,550 | 0.003 | 0.999 |
| Unemployment / saving rates | 6,550 | 0.014 | 0.993 |
| Unemployment / stock prices | 6,550 | 0.022 | 0.989 |
| Saving rates / stock prices | 6,550 | 0.029 | 0.986 |

*Note:* Rows use the exact-12 raw-Pearson common-support sample. Focal response controls are the respondent-level shares of 0, 50, and 100 answers in the three probability inputs and the overall focal share. The table assesses whether primitive correlations are mostly absorbed by focal probability-scale use.

Table C.14 adds a broader probability-format diagnostic. It compares the primary macro probability-comovement index to auxiliary probability questions that are not part of the benchmark macro block. The move-residence and debt-payment placebo pair is available for most exact-12 respondents and is the primary non-macro check in the table. The personal-labor rows are more selected because the SCE routes them to eligible respondents, but they serve as secondary checks because they use the same 0–100 probability scale in another domain. The table also reports simple auxiliary probability-answer summaries, including calendar-adjacent month-to-month auxiliary revisions. These checks test whether the primary macro probability ranking is mostly a generic probability-answering tendency.

Table C.14: Auxiliary probability diagnostics for the primary SCE measure

| Diagnostic | $N$ | Mean | p5 | Median | p95 | Corr. with primary |
|---|---|---|---|---|---|---|
| Move/debt placebo correlation | 5,005 | 0.106 | -0.456 | 0.083 | 0.747 | 0.042 |
| Job loss/quit correlation | 2,834 | 0.252 | -0.418 | 0.270 | 0.866 | 0.082 |
| Personal labor probability index | 2,808 | 0.097 | -0.259 | 0.074 | 0.529 | 0.106 |
| Auxiliary probability mean | 6,550 | 18.573 | 0.167 | 17.505 | 45.094 | -0.051 |
| Auxiliary probability answer SD | 6,550 | 20.592 | 0.451 | 21.579 | 39.423 | -0.039 |
| Auxiliary focal-answer share | 6,550 | 0.377 | 0.017 | 0.300 | 0.958 | 0.050 |
| Auxiliary average absolute month-to-month revision | 6,550 | 7.534 | 0.182 | 6.709 | 18.136 | 0.083 |

*Note:* The first three rows are within-respondent auxiliary probability-comovement measures computed over the exact-12 common-support sample when the listed auxiliary series are complete and have nonzero within-person variance. The remaining rows summarize auxiliary probability answers from moving, debt-payment, and routed personal-labor probability questions; the revision row uses only calendar-adjacent month-to-month changes. The final column reports the across-respondent correlation with the primary macro probability signed-comovement index.

## C.7 Sample Robustness

Table C.15 compares the exact-12 benchmark with two alternative sample-window choices. The consecutive exact-12 row restricts the benchmark to respondents whose 12 complete waves are calendar-consecutive. The at-least-12 row keeps respondents with at least 12 complete waves and uses their first 12 complete waves. All rows apply the common-support restriction, so their respondent counts are below the corresponding pre-common-support rows of Table C.1. The exact-12 design remains the benchmark for the reasons given in Section 3.4; the main probability-versus-inflation ranking is not driven by the sample-window choice.

Table C.15: SCE sample-window robustness for signed-comovement means

| Sample | *N* | Probability | Inflation | Overall | Gap |
|---|---|---|---|---|---|
| Exact-12 benchmark | 6,550 | 0.222 | 0.033 | 0.128 | 0.189 |
| Consecutive exact-12 | 6,480 | 0.222 | 0.033 | 0.127 | 0.189 |
| At-least-12, first 12 | 6,980 | 0.220 | 0.033 | 0.127 | 0.187 |

*Note:* Rows use raw Pearson signed-comovement measures on common-support respondents, so the counts are below the corresponding pre-common-support rows of the sample-construction table. The benchmark exact-12 sample is used in the main text because it gives a common panel length and a transparent split-half design; nearby sample-window choices preserve the main probability-versus-inflation ranking.

Table C.16 tests whether the exact-12 design is fragile because it leaves longer respondent histories unused. The public SCE contains relatively few respondents with more than 12 complete benchmark waves. After common-support restrictions, 430 respondents have more than 12 complete waves and the maximum retained history is 16 waves. The all-available at-least-12 row changes neither the probability-versus-inflation ranking nor the reliability hierarchy. The longer-only row is a diagnostic only: it is too small and too selected to replace the exact-12 benchmark.

Table C.16: SCE panel-length audit

| Sample | *N* | Mean waves | Max waves | Prob. mean | Infl. mean | Prob. rel. | Infl. rel. |
|---|---|---|---|---|---|---|---|
| Exact-12 benchmark | 6,550 | 12.00 | 12 | 0.222 | 0.033 | 0.394 | 0.114 |
| At-least-12, first 12 | 6,980 | 12.00 | 12 | 0.220 | 0.033 | 0.394 | 0.116 |
| At-least-12, all available | 6,980 | 12.08 | 16 | 0.220 | 0.034 | 0.395 | 0.118 |
| More-than-12, all available | 430 | 13.26 | 16 | 0.189 | 0.041 | 0.403 | 0.182 |

*Note:* Rows use raw Pearson signed-comovement measures on common-support respondents. The all-available rows use every complete benchmark-input wave retained by the row's sample rule. Reliability is the mean across 200 balanced random splits of each respondent's available waves; odd panel lengths are split into floor and ceiling halves. The more-than-12 row is diagnostic because the public SCE contains little usable panel depth beyond 12 complete waves.

Table C.17 summarizes retained and excluded respondents by benchmark-panel status. The table documents the measurement-sample restriction.

Table C.17: Retained and excluded SCE respondents

| Group | Respondents | Complete waves | Focal share | Spending tail | Infl. | Unemp. | Saving | Stocks |
|---|---|---|---|---|---|---|---|---|
| Exact-12 benchmark | 6,659 | 12.0 | 18.8 | 0.81 | 4.09 | 37.68 | 33.05 | 42.27 |
| Fewer than 12 complete waves | 16,422 | 5.5 | 19.3 | 1.57 | 4.23 | 39.71 | 33.02 | 43.29 |
| More than 12 complete waves | 437 | 13.3 | 18.1 | 0.47 | 5.02 | 38.68 | 32.84 | 41.41 |
| Zero complete waves | 368 | 0.0 | 20.1 | 4.03 | 4.68 | 41.62 | 34.57 | 43.44 |

*Note:* Groups are based on complete observations for the four benchmark measurement inputs. Focal share is the respondent-level share, in percent, of probability-input answers equal to 0, 50, or 100. Spending tail is the respondent-level share, in percent, of nonmissing expected-spending-growth observations with absolute value above 50 percentage points. Input means are respondent-level averages, then averaged within group.

## C.8 Additional SCE Validation

Table C.18 reports the time-permutation null separately for each primitive correlation. The main text reports the reduced-index nulls; this table shows that the probability-block result is present at the primitive-correlation level.

Table C.18: Permutation null benchmark for mean correlations

| Correlation pair | Observed | Null p5 | Null p50 | Null p95 |
|---|---|---|---|---|
| Inflation / unemployment | 0.065 | -0.005 | 0.000 | 0.006 |
| Inflation / saving rates | 0.041 | -0.005 | 0.000 | 0.007 |
| Inflation / stock prices | -0.007 | -0.007 | 0.000 | 0.007 |
| Unemployment / saving rates | 0.221 | -0.006 | -0.001 | 0.004 |
| Unemployment / stock prices | 0.150 | -0.005 | 0.000 | 0.007 |
| Saving rates / stock prices | 0.295 | -0.007 | -0.001 | 0.007 |

*Note:* The null benchmark uses 100 within-respondent, variable-specific time permutations. The table compares the observed mean raw-Pearson correlation to the empirical distribution of null mean correlations.

Table C.19 splits exact-12 respondents by the midpoint of their retained measurement window. The table checks whether the main probability-versus-inflation ranking exists only in one macro period.

Table C.19: SCE timing-window diagnostic

| Window | Respondents | First | Last | Probability | Inflation | Overall | Gap |
|---|---|---|---|---|---|---|---|
| All exact-12 common support | 6,550 | 2013m6 | 2025m4 | 0.222 | 0.033 | 0.128 | 0.189 |
| Pre-pandemic midpoint | 3,716 | 2013m6 | 2020m7 | 0.228 | 0.031 | 0.129 | 0.197 |
| Pandemic/reopening midpoint | 1,176 | 2019m9 | 2022m5 | 0.174 | 0.011 | 0.093 | 0.163 |
| 2022-onward midpoint | 1,658 | 2021m7 | 2025m4 | 0.243 | 0.053 | 0.148 | 0.190 |

*Note:* Rows use exact-12 raw-Pearson common-support respondents. Timing windows are assigned by the median month-serial value of the respondent's retained 12-wave measurement window. The check asks whether the probability-versus-inflation ranking appears only in one macro period.

Table C.20 adds the jackknife diagnostic to the reliability hierarchy reported in the main text. The jackknife check is less demanding than split-half validation because leave-one-wave-out scores overlap with the full-12 score, but it checks mechanically whether a single survey month drives the respondent ordering.

Table C.20: Reliability of reduced-dimension comovement measures

| Measure | Random | SB | First/last | Odd/even | Jackknife min |
|---|---|---|---|---|---|
| Probability | 0.394 | 0.565 | 0.283 | 0.436 | 0.945 |
| Inflation | 0.114 | 0.204 | 0.011 | 0.152 | 0.860 |
| Overall | 0.277 | 0.433 | 0.174 | 0.317 | 0.909 |
| Probability − inflation | 0.255 | 0.406 | 0.148 | 0.294 | 0.914 |
| Projected PC1 | 0.369 | 0.539 | 0.259 | 0.412 | 0.936 |
| Projected PC2 | 0.133 | 0.234 | 0.026 | 0.173 | 0.871 |

*Note:* Random-split statistics average 200 respondent-specific random 6/6 splits. SB is the Spearman-Brown adjustment. The jackknife column reports the minimum full-12 versus leave-one-wave-out correlation across the 12 omitted waves and is less demanding because 11 waves overlap.

Table C.21 translates the split-half reliability of the primary index into ranking terms. Under classical measurement error with normal true scores and noise, and with the reliability set to the Spearman-Brown 12-wave value of 0.565, the simulation reports where respondents in each observed-index quartile sit in the true-score distribution. Observed top-quartile respondents have an expected true-score percentile of 0.78, are in the true top quartile with probability 0.64, and fall below the true median about 10 percent of the time; the bottom quartile is the mirror image. The calibration quantifies the rank uncertainty implied by the reported reliability. All regression and correlation results in the paper use the observed index, so under classical assumptions this noise attenuates rather than overstates the reported associations.

Table C.21: True-score rank calibration for the probability signed-comovement index

| Observed quartile | Expected true percentile | Same quartile | True top quartile | True below median |
|---|---|---|---|---|
| Bottom quartile | 0.224 | 0.638 | 0.012 | 0.900 |
| Second quartile | 0.420 | 0.379 | 0.087 | 0.642 |
| Third quartile | 0.581 | 0.380 | 0.262 | 0.358 |
| Top quartile | 0.777 | 0.639 | 0.639 | 0.100 |

*Note:* Simulation under classical measurement error with normal true scores and noise: the observed index equals the square root of the reliability times a standard-normal true score plus independent noise, with the reliability set to the Spearman-Brown 12-wave value of 0.565 for the probability signed-comovement index. Rows condition on the observed-index quartile; columns report the expected percentile of the true score and the probabilities that the true score falls in the same quartile, in the top quartile, and below the median. Based on 2,000,000 simulation draws with a fixed seed. All regression and correlation results in the paper use the observed index, so the table quantifies rank attenuation; it does not alter the reported associations.

Table C.22 reports a within-system revision diagnostic for the SCE probability block. The diagnostic uses adjacent calendar-consecutive transitions in the exact-12 common-support sample and asks whether pairs of probability expectations move in the same direction when both answers change. The null independently permutes each probability input's adjacent revisions within respondent, preserving marginal revision behavior while breaking within-transition alignment.

Table C.22: Within-system probability revision diagnostic

| Quartile | Respondents | Pairs | Index | Same dir. | Null median | Abs. rev. |
|---|---|---|---|---|---|---|
| Q1 low | 1,638 | 35,626 | -0.151 | 0.462 | 0.498 | 13.37 |
| Q2 | 1,637 | 36,694 | 0.091 | 0.559 | 0.501 | 14.45 |
| Q3 | 1,637 | 38,263 | 0.313 | 0.635 | 0.503 | 14.99 |
| Q4 high | 1,638 | 39,307 | 0.635 | 0.737 | 0.509 | 14.58 |

*Note:* Respondents are sorted into quartiles of the exact-12 probability signed-comovement index. Revision pairs are adjacent, calendar-consecutive transitions for the three probability inputs where both variables in a pair change. Same dir. is the respondent-weighted share of active revision pairs that move in the same direction. The null independently permutes each probability input's adjacent revisions within respondent over 100 iterations, preserving each respondent's marginal revision distribution while breaking within-transition alignment. Abs. rev. is the respondent-level mean absolute probability revision in percentage points.

Table C.23 reports descriptive projections of the reduced SCE measures on expectation levels, response-style measures, timing variables, demographics, education, income, and numeracy.

Table C.23: Descriptive reducibility to observed controls

| Measure | Levels | Response | Timing | Dem./num. | All |
|---|---|---|---|---|---|
| Probability | 0.215 | 0.032 | 0.000 | 0.142 | 0.286 |
| Inflation | 0.007 | 0.002 | 0.001 | 0.002 | 0.012 |
| Overall | 0.155 | 0.023 | 0.001 | 0.086 | 0.194 |
| Probability − inflation | 0.107 | 0.017 | 0.001 | 0.088 | 0.160 |

*Note:* Entries are R-squared values from descriptive linear projections of each comovement measure on the listed control block. The diagnostics assess reducibility.

Table C.24 complements the block R-squared values with signed gradients for the primary measure. The R-squared entries say how much of probability signed comovement the observed controls explain; this table reports which respondent characteristics are associated with higher comovement. Probability comovement is higher among female respondents and lower among respondents with higher numeracy, college education, and higher household income, while age and region gradients are small. The demographic and numeracy gradients shrink but do not disappear when average expectation levels and the focal-answer share are added in column (2). The large unexplained residual from Table C.23 applies here as well.

Table C.25 reports the uncertainty diagnostic flagged in Section 3.6. The SCE inflation-density question gives one observed proxy for broader perceived macro uncertainty: the density interquartile range. This proxy is relevant because mean reported probabilities are below 50 for all three inputs, so greater perceived uncertainty can compress directional probability answers toward 50 and raise all three together.

Panels A and C show the cross-sectional association. Respondents with a higher average density IQR rank higher on the probability signed-comovement index, with a raw correlation of 0.234. The corresponding OLS association is 0.027 index points per cross-sectional standard deviation of average IQR under demographic and numeracy controls, and 0.026 after adding

Table C.24: Descriptive correlates of probability signed comovement

| Correlate | (1) | (2) |
|---|---|---|
| Female | 0.115*** | 0.078*** |
| | (0.007) | (0.007) |
| Age 40 to 60 | -0.006 | -0.018** |
| | (0.009) | (0.008) |
| Age over 60 | -0.002 | -0.018** |
| | (0.009) | (0.009) |
| Some college | -0.038*** | -0.020 |
| | (0.013) | (0.012) |
| College | -0.081*** | -0.042*** |
| | (0.013) | (0.012) |
| Income 50k to 100k | -0.031*** | -0.038*** |
| | (0.009) | (0.008) |
| Income over 100k | -0.048*** | -0.048*** |
| | (0.010) | (0.009) |
| Numeracy share (0–1) | -0.297*** | -0.185*** |
| | (0.019) | (0.018) |
| Mean inflation expectation (per SD) | – | -0.013*** |
| | | (0.004) |
| Mean Pr(unemployment higher) (per SD) | – | -0.062*** |
| | | (0.004) |
| Mean Pr(saving rates higher) (per SD) | – | -0.007** |
| | | (0.004) |
| Mean Pr(stock prices higher) (per SD) | – | -0.083*** |
| | | (0.004) |
| Focal-answer share (0–1) | – | -0.043* |
| | | (0.023) |
| Observations | 6,550 | 6,550 |
| R-squared | 0.142 | 0.267 |

*Note:* The dependent variable is the probability signed-comovement index in the exact-12 raw-Pearson common-support sample (mean 0.222, standard deviation 0.305). Entries are OLS coefficients with heteroskedasticity-robust (HC1) standard errors in parentheses. Column (1) regresses the index on demographics and numeracy; column (2) adds the four average expectation levels, standardized to one cross-sectional standard deviation each, and the overall focal-answer share. Reference categories are male, age under 40, high-school education, and household income under 50k. Region indicators, missing-category indicators, and auxiliary numeracy controls are included in both columns but not shown. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

expectation levels and the focal-answer share.

Panel B shows the within-person version. The mean correlation between the density IQR and the three probability inputs is 0.053, above a within-person time-permutation null whose 95th percentile is 0.004, and the link rises monotonically across quartiles of the primary index, from 0.004 in the bottom quartile to 0.134 in the top quartile. Because common months could also generate a positive raw link, Panel B repeats the test with month-standardized inputs. The link is nearly unchanged, at 0.050 against a null 95th percentile of 0.005, with a nearly identical quartile gradient from −0.002 to 0.132.

The diagnostic uses inflation uncertainty as a proxy for broader perceived macro uncertainty, so it is supporting evidence rather than a full mechanism test. Panel D quantifies that limit: residualizing each probability input on the within-person IQR and recomputing the index from the residual series leaves the mean at 0.219, versus 0.222 for the same 6,535 respondents, and the residualized and unresidualized rankings correlate at 0.959. The index is not a measured-inflation-uncertainty index; instead, the IQR evidence shows that households who appear more uncertain about inflation also organize other macro probabilities more tightly, while the broader uncertainty state is only partly observed.

Table C.25: Inflation-density uncertainty and probability signed comovement

| Quantity | $N$ | Mean | p5 | Median | p95 | Corr. with primary |
|---|---|---|---|---|---|---|
| *Panel A: Respondent-average inflation uncertainty* | | | | | | |
| Mean inflation-density IQR (percentage points) | 6,550 | 4.054 | 1.095 | 2.922 | 11.468 | 0.234 |
| *Panel B: Within-person uncertainty-link index* | | | | | | |
| Uncertainty-link index | 6,538 | 0.053 | -0.336 | 0.040 | 0.490 | 0.213 |
| Within-person time-permutation null | 6,538 | 0.000 | – | – | 0.004 | – |
| Link index, quartile 1 of primary index | 1,635 | 0.004 | – | – | – | – |
| Link index, quartile 2 of primary index | 1,634 | 0.023 | – | – | – | – |
| Link index, quartile 3 of primary index | 1,634 | 0.052 | – | – | – | – |
| Link index, quartile 4 of primary index | 1,635 | 0.134 | – | – | – | – |
| Uncertainty-link index, month-standardized inputs | 6,550 | 0.050 | -0.333 | 0.037 | 0.492 | 0.216 |
| Within-person time-permutation null | 6,550 | 0.000 | – | – | 0.005 | – |
| Month-standardized link index, quartile 1 of primary index | 1,638 | -0.002 | – | – | – | – |
| Month-standardized link index, quartile 2 of primary index | 1,637 | 0.023 | – | – | – | – |
| Month-standardized link index, quartile 3 of primary index | 1,637 | 0.047 | – | – | – | – |
| Month-standardized link index, quartile 4 of primary index | 1,638 | 0.132 | – | – | – | – |
| *Panel C: OLS of primary index on mean IQR (per cross-sectional SD)* | | | | | | |
| Demographics and numeracy controls | 6,550 | $0.027^{***}$ (0.004), $R^2 = 0.148$ | | | | |
| Plus expectation levels and focal share | 6,550 | $0.026^{***}$ (0.004), $R^2 = 0.273$ | | | | |
| *Panel D: IQR-residualized probability comovement* | | | | | | |
| Probability index, IQR-residualized inputs | 6,535 | 0.219 | -0.237 | 0.197 | 0.747 | 0.959 |
| Unresidualized index, same respondents | 6,535 | 0.222 | -0.233 | 0.198 | 0.764 | – |

*Note:* Panel A summarizes each respondent's average density-implied inflation interquartile range over the 12 retained waves, computed for exact-12 common-support respondents with 12 nonmissing values. Panel B summarizes the uncertainty-link index, the mean within-person correlation between the inflation-density IQR and the three probability inputs, computed when all four series are complete and have nonzero within-person variance; the null row reports the mean and 95th percentile of the cross-respondent average after permuting each respondent's IQR series over waves, and the quartile rows report the mean link index by quartile of the primary probability signed-comovement index. The month-standardized rows recompute the link index after removing common survey-month location and scale from both the IQR and the probability inputs, using the same survey-month standardization as the month-standardized variant of the primary index. The final column is the across-respondent correlation with the primary index. Panel C reports the OLS coefficient on the mean IQR, standardized to one cross-sectional standard deviation, under the two respondent-characteristic control specifications with HC1 standard errors in parentheses. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Panel D residualizes each probability input on the inflation-density IQR within respondent over the 12 waves and recomputes the probability signed-comovement index from the residual series; the comparison row reports the unresidualized index for the same respondents. The IQR measures inflation uncertainty only and serves as a proxy for broader perceived macro uncertainty.

## C.9 Additional ECB CES Evidence

Table C.26 reports the six primitive raw-Pearson correlations behind the ECB adverse macro-comovement index. The table shows that the ECB evidence is more than a single average index. The unemployment/economy-weakness link is the largest primitive component, while the other links are also centered above zero. This pattern is consistent with broad adverse macro comovement.

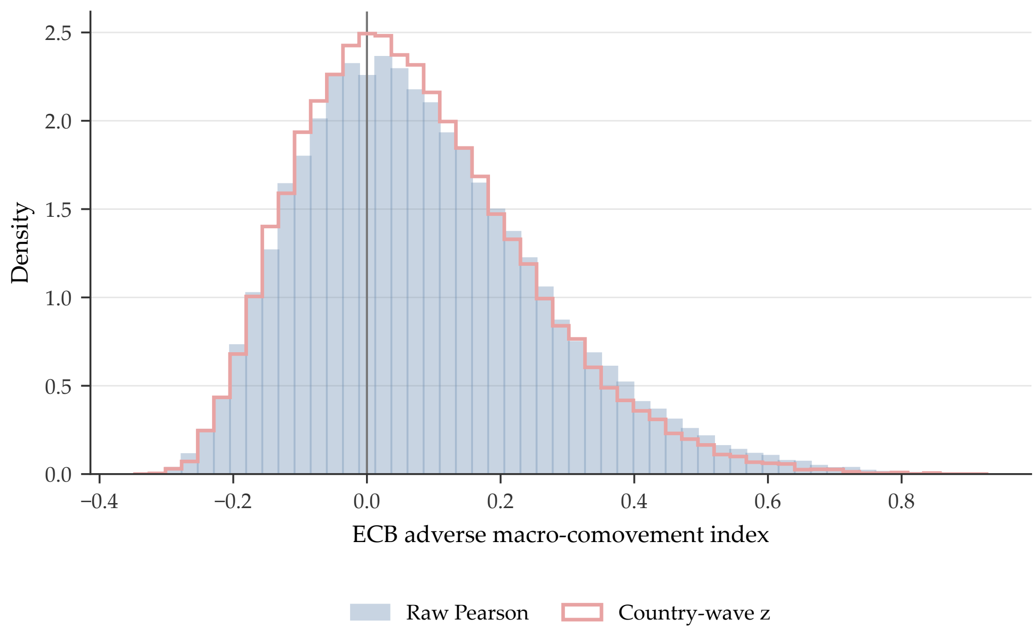
Table C.26: ECB CES primitive adverse-comovement correlations

| Primitive correlation | *N* | Mean | Median | p5 | p95 |
|---|---|---|---|---|---|
| Inflation / unemployment | 38,885 | 0.053 | 0.055 | -0.532 | 0.626 |
| Inflation / mortgage rates | 38,885 | 0.073 | 0.074 | -0.526 | 0.665 |
| Inflation / economy weakness | 38,885 | 0.095 | 0.107 | -0.565 | 0.713 |
| Unemployment / mortgage rates | 38,885 | 0.038 | 0.041 | -0.543 | 0.607 |
| Unemployment / economy weakness | 38,885 | 0.191 | 0.207 | -0.435 | 0.757 |
| Mortgage rates / economy weakness | 38,885 | 0.051 | 0.055 | -0.539 | 0.633 |

*Note:* Rows report raw-Pearson within-respondent correlations for the first-12 ECB comparison sample. The signs follow the adverse macro-comovement convention used in the main ECB index.

Figure C.3 plots the raw and standardized ECB adverse-comovement distributions. The standardized version has a similar center and range, so the result is not just a country-wave artifact.

Figure C.3: ECB CES adverse macro-comovement distributions




*Note.* The figure plots density-normalized histograms of the ECB adverse macro-comovement index. The raw-Pearson benchmark is shown as a filled distribution; the line shows the country-wave-standardized Pearson variant. Histogram ranges include all observed values for the plotted measures.

Table C.27 checks whether the ECB comparison depends on using the open-ended inflation expectation rather than a density-style inflation measure closer to the SCE benchmark. The robustness row replaces open-ended inflation with a midpoint-implied mean from the CES probabilistic inflation bins after strict bin validation. The adverse-comovement mean is nearly identical, 0.085 versus 0.082, so the cross-survey result does not depend on the open-ended inflation elicitation.

Table C.27: ECB CES inflation-elicitation robustness

| Inflation input | Sample | $N$ | Overall mean | Overall median | Inflation-link mean | Noninflation mean |
|---|---|---|---|---|---|---|
| Open-ended inflation | Headline first-12 sample | 38,885 | 0.084 | 0.060 | 0.074 | 0.093 |
| Open-ended inflation | Density-valid first-12 sample | 38,418 | 0.085 | 0.061 | 0.076 | 0.093 |
| Density-bin inflation | Density-valid first-12 sample | 38,370 | 0.082 | 0.060 | 0.072 | 0.093 |

*Note:* The headline ECB row uses the open-ended inflation expectation. The density-valid rows use the first 12 complete observations among respondents with valid probabilistic inflation bins; the second row keeps open-ended inflation on that sample and the third row replaces it with a midpoint-implied mean from the ECB probabilistic bins. Valid bin rows require complete bins in the 0–100 range that sum to 100. All rows report raw-Pearson adverse macro-comovement measures.

Table C.28 reports the raw-Pearson ECB adverse macro-comovement index by country as a concentration check: it tests whether the cross-survey result is confined to a single national sample. Country samples differ in macroeconomic conditions, fielding periods, response behavior, and the interpretation of quantitative expectations, so the rows should not be read as a ranking of countries.

Table C.28: ECB CES adverse macro-comovement by country

| Country | $N$ | Mean | Median | SD |
|---|---|---|---|---|
| AT | 1,500 | 0.075 | 0.052 | 0.167 |
| BE | 2,072 | 0.080 | 0.058 | 0.177 |
| DE | 6,580 | 0.080 | 0.053 | 0.184 |
| EL | 1,280 | 0.079 | 0.067 | 0.167 |
| ES | 6,929 | 0.076 | 0.053 | 0.179 |
| FI | 1,616 | 0.089 | 0.067 | 0.176 |
| FR | 6,513 | 0.089 | 0.064 | 0.179 |
| IE | 1,293 | 0.091 | 0.071 | 0.180 |
| IT | 7,170 | 0.093 | 0.069 | 0.183 |
| NL | 2,195 | 0.075 | 0.054 | 0.173 |
| PT | 1,737 | 0.085 | 0.055 | 0.182 |

*Note:* Rows report the raw-Pearson adverse macro-comovement index by country in the ECB CES first-12 comparison sample. Country rows are concentration diagnostics for the cross-survey result.

Table C.29 reports descriptive R-squared values from projections of the ECB adverse macro-comovement index on broad country and timing indicators. The explanatory power is small, including when country indicators are combined with entry-cohort or window-midpoint cohort indicators. The ECB result is not a broad geography-timing cell artifact.

Table C.29: ECB CES timing and country explanatory power

| Design | $N$ | $R^2$ |
|---|---|---|
| Entry cohort | 38,885 | 0.000 |
| Country | 38,885 | 0.001 |
| Country and entry cohort | 38,885 | 0.001 |
| Country and window-midpoint cohort | 38,885 | 0.011 |

*Note:* Entries are R-squared values from descriptive projections of the raw-Pearson ECB adverse macro-comovement index on the listed country and timing indicators.

These diagnostics underlie the cross-survey conclusion in Section 3.7: within-person comovement of adverse macro expectations appears in another public expectations panel and holds under simple country, timing, tail, and scale checks, with the limits on respondent-level reliability stated there.

## C.10 Exploratory Evidence on Discrete Types

A natural alternative is that household belief systems can be represented by a small number of discrete types. Table C.30 reports clustering diagnostics for the raw-Pearson six-correlation matrix. The fixed full-sample solutions are stable across random starts and bootstrap resamples, but split-half adjusted Rand indices are near zero when classes are estimated separately from the first six and last six waves. Continuous signed-comovement measures are more stable and easier to interpret.

Table C.30: Exploratory SCE clustering diagnostics

| $k$ | $N$ | Silhouette | Seed ARI | Bootstrap ARI | Split-half ARI | Min. share | Max. share |
|---|---|---|---|---|---|---|---|
| 2 | 6,550 | 0.189 | 0.996 | 0.964 | 0.036 | 44.0 | 56.0 |
| 3 | 6,550 | 0.178 | 0.999 | 0.953 | 0.027 | 26.1 | 46.8 |
| 4 | 6,550 | 0.180 | 0.983 | 0.895 | 0.008 | 22.5 | 28.4 |
| 5 | 6,550 | 0.193 | 0.986 | 0.909 | 0.008 | 15.9 | 22.7 |
| 6 | 6,550 | 0.190 | 0.987 | 0.918 | 0.008 | 14.2 | 21.9 |

*Note:* Rows use the raw-Pearson six-feature matrix in the exact-12 common-support sample. Seed ARI compares labels from alternative random starts with the baseline full-sample solution; bootstrap ARI compares baseline labels with models fitted on bootstrap resamples and predicted for all respondents. Split-half ARI clusters first-six and last-six features separately and compares respondent labels across halves. The low split-half ARI is the main reason discrete classes are kept exploratory.

The clustering results are still informative. Stable full-sample and bootstrap partitions show that the six-correlation space has stable full-sample structure rather than unstructured dispersion. The split-half adjusted Rand indices are the harder test, because a solution can look interpretable in the full sample while the first six and last six waves imply unrelated labels for the same respondents. In that case class names would mainly describe the fitted full-sample partition rather than durable household categories, and strong labels would overstate the stability of discrete groups.

The continuous probability signed-comovement index answers the central question about types in the form the public SCE can support: households differ persistently in the degree to which probability expectations move together. Future work with longer panels, targeted subjective-model modules, or experimental designs could revisit discrete types; here, discrete classes remain a descriptive supplement and the continuous ranking is the main measure.

## C.11 Additional Prediction and Behavior Checks

The main text reports the expectation-dynamics and spending-slope specifications and summarizes the direct spending benchmark. Table C.31 documents the spending sample and tail rule. Table C.32 reports the direct spending-level specifications summarized in the main text. Table C.33 reports secondary measurement regressors in the same respondent-level full-control specification used for the main trimmed spending result.

Table C.31: Spending analysis sample

| Metric | Value |
|---|---|
| Exact-12 respondents | 6,659 |
| Exact-12 respondent-months | 79,908 |
| Common-support respondents | 6,550 |
| Respondents merged with spending data | 6,550 |
| Respondent-months merged with spending data | 78,600 |
| Respondent-months with spending expectations | 78,562 |
| Respondent-months with absolute spending growth above 50 pp | 605 |
| Respondent-months in primary spending outcome | 77,957 |
| Share above tail cutoff (percent) | 0.77 |
| Respondents with primary outcome | 6,549 |
| Mean primary spending outcome | 4.360 |
| Mean winsorized spending outcome | 4.540 |

*Note:* The behavioral sample starts from the exact-12 measurement panel and keeps respondents in the raw-Pearson common-support feature file. The primary spending outcome drops respondent-month observations with absolute expected spending growth above 50 percentage points.

Measurement error cannot make the direct spending null economically large. Because the regressor is standardized, classical measurement error attenuates the coefficient by the square root of the index's reliability: with the adjusted 12-wave reliability of 0.565, the disattenuated respondent-level coefficient is about −0.02 percentage points per standard deviation of true comovement, and about −0.03 under the more conservative six-wave split reliability of 0.394. Even the boundary of the disattenuated 95 percent confidence interval stays below 0.35 percentage points in absolute value, less than a twentieth of the outcome standard deviation of

Table C.32: Reduced-form spending relevance of the primary SCE comovement measure

| Variable | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Probability comovement | -0.017 | -0.009 | -0.095 | -0.105 | -0.046 | -0.019 |
| | (0.102) | (0.106) | (0.097) | (0.106) | (0.106) | (0.109) |
| Design | Respondent | Respondent | Panel | Panel | First half → later | First half → later |
| Outcome | Trimmed | Winsorized | Trimmed | Winsorized | Trimmed | Winsorized |
| Month fixed effects | – | – | Yes | Yes | – | – |
| Full controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 6,549 | 6,550 | 77,957 | 78,562 | 6,419 | 6,421 |
| $R^2$ | 0.072 | 0.077 | 0.040 | 0.044 | 0.047 | 0.052 |
| Mean dep. var. | 4.360 | 4.540 | 4.318 | 4.540 | 3.932 | 4.109 |

*Note:* The dependent variable is expected spending growth in percentage points. The regressor is the primary SCE probability signed-comovement index, standardized to one cross-sectional standard deviation in each estimation sample; standard errors are in parentheses. All columns include the full control block: expectation levels, response-style and timing measures, demographics, education, income, and numeracy. Respondent columns regress respondent-average expected spending growth on the index, with heteroskedasticity-robust (HC1) standard errors. Panel columns use respondent-month observations with survey-month fixed effects and standard errors clustered at the respondent level. First-half columns predict average wave-7–12 expected spending growth from wave-1–6 comovement, with HC1 standard errors. Trimmed outcomes drop observations with absolute expected spending growth above 50 percentage points; winsorized outcomes apply month-level winsorization. None of the estimates are causal effects. Significance levels: $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

7.2 percentage points.

Table C.33: Secondary spending regressors

| Regressor | Coef. | SE | $N$ | $R^2$ |
|---|---|---|---|---|
| Probability signed comovement | -0.017 | (0.102) | 6,549 | 0.072 |
| Inflation signed comovement | -0.117 | (0.088) | 6,549 | 0.072 |
| Overall signed comovement | -0.098 | (0.092) | 6,549 | 0.072 |
| Probability minus inflation | 0.078 | (0.099) | 6,549 | 0.072 |
| PC1 | -0.044 | (0.098) | 6,549 | 0.072 |
| PC2 | -0.109 | (0.090) | 6,549 | 0.072 |

*Note:* Each row is a separate respondent-level regression of the primary trimmed spending outcome on one standardized measurement regressor and the full control set, with heteroskedasticity-robust (HC1) standard errors in parentheses. Primitive-correlation regressions are similar and omitted for brevity. Significance levels: $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.34 reports the respondent-month panel specifications behind the summary behavioral table in the main text. These rows use survey-month fixed effects and respondent-clustered standard errors. They compare respondents within the same survey month while retaining the repeated spending-expectation observations. The coefficients remain small relative to their standard errors, so the panel design does not change the interpretation.

Tables C.35 and C.36 report the timing-separated spending exercise. The regressor is recomputed from waves 1–6 and the outcome is average expected spending growth in waves 7–12. This separation makes the timing interpretation cleaner than the contemporaneous

Table C.34: Panel robustness for spending

| Specification | Coef. | SE | $N$ | Clusters | $R^2$ |
|---|---|---|---|---|---|
| Panel trimmed | -0.095 | (0.097) | 77,957 | 6,549 | 0.040 |
| Panel winsorized | -0.105 | (0.106) | 78,562 | 6,550 | 0.044 |

*Note:* Panel specifications use respondent-month observations, survey-month fixed effects, full respondent-level controls, and respondent-clustered standard errors. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

respondent-level specification, but it also makes the measure noisier because each respondent's comovement index is estimated from only six observations.

Table C.35: First-half to second-half spending sample

| Metric | Value |
|---|---|
| Respondents with first-half measures | 6,659 |
| Non-missing first-half probability comovement | 6,421 |
| Respondents with later spending outcome | 6,657 |
| Primary estimation respondents | 6,419 |
| Wave 7–12 respondent-months with spending expectations | 39,934 |
| Wave 7–12 respondent-months with absolute spending growth above 50 pp | 276 |
| Wave 7–12 respondent-months in primary outcome | 39,658 |
| Share above tail cutoff, waves 7–12 (percent) | 0.69 |
| Mean primary outcome (waves 7–12) | 4.005 |
| Mean winsorized outcome (waves 7–12) | 4.183 |

*Note:* The timing robustness computes raw-Pearson comovement using waves 1–6 and predicts average expected spending growth in waves 7–12. This timing separation avoids contemporaneous overlap but increases measurement noise because the comovement measure uses only six observations.

Across the respondent-level, panel, secondary-regressor, and first-half-to-later-spending rows, the estimated spending associations are economically small and imprecise.

**Additional predictive-content checks**

The spending-slope interaction motivated by Equation (3.1) is reported in the main text (Section 3.8.3, Table 3.7). This appendix adds robustness checks for that interaction and the secondary personal-risk expectations.

Table C.37 reports variants of the spending-slope interaction. The uncertainty rows address the closest competing interpretation: because the probability index correlates with average inflation uncertainty, the slope heterogeneity could in principle reflect an uncertainty-inflation interaction. Controlling for average inflation uncertainty leaves the interaction negative and significant. In the joint specification with both comovement-inflation and uncertainty-inflation interactions, the comovement interaction remains the larger term while the uncertainty interaction is smaller and statistically insignificant. The joint specification cannot cleanly separate the two correlated channels, but average inflation uncertainty does not account for the

Table C.36: First-half comovement and future spending

| Specification | Coef. | SE | $N$ | $R^2$ |
|---|---|---|---|---|
| No controls | -0.073 | (0.097) | 6,419 | 0.000 |
| First-half expectation levels | -0.022 | (0.104) | 6,419 | 0.032 |
| Full first-half controls | -0.046 | (0.106) | 6,419 | 0.047 |
| Winsorized future outcome | -0.019 | (0.109) | 6,421 | 0.052 |

*Note:* The coefficient is percentage points of average wave-7–12 expected spending growth for a one-standard-deviation increase in probability signed comovement measured over waves 1–6, with heteroskedasticity-robust (HC1) standard errors in parentheses. All controls in the controlled specifications are measured from waves 1–6, except demographics and numeracy. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

interaction.

The numeracy-interaction rows address the related reading that high comovement proxies for less sophisticated probabilistic answering. With a numeracy-inflation interaction included, the comovement interaction remains negative and significant while the numeracy interaction is positive and marginally significant. Its sign is consistent with some role for sophistication, but it does not absorb the comovement interaction. The timing rows show that the interaction is concentrated in measurement windows centered on 2022 onward and is small and insignificant in earlier windows. The inflation-level columns show that the baseline inflation–spending slope is positive and significant in every window and steepest from 2022 onward. The point-forecast row attenuates toward zero, as expected when the interacted inflation level is replaced by a noisier elicitation. The panel row reports the same interaction within survey months with respondent-clustered standard errors.

Table C.37: Spending-slope interaction robustness

| Specification | Interaction | SE | Infl. level | SE | $N$ | $R^2$ |
|---|---|---|---|---|---|---|
| Baseline, trimmed outcome | -0.316*** | (0.115) | 1.535*** | (0.136) | 6,549 | 0.073 |
| Baseline, winsorized outcome | -0.381*** | (0.118) | 1.585*** | (0.136) | 6,550 | 0.080 |
| Average inflation-uncertainty control | -0.260** | (0.114) | 1.439*** | (0.137) | 6,549 | 0.078 |
| Joint specification: comovement × inflation | -0.219* | (0.120) | 1.630*** | (0.154) | 6,549 | 0.079 |
| Joint specification: uncertainty × inflation | -0.172 | (0.114) | 1.630*** | (0.154) | 6,549 | 0.079 |
| With numeracy interaction: comovement × inflation | -0.269** | (0.119) | 1.628*** | (0.130) | 6,548 | 0.074 |
| With numeracy interaction: numeracy × inflation | 0.196* | (0.113) | 1.628*** | (0.130) | 6,548 | 0.074 |
| Pre-pandemic midpoint | -0.142 | (0.158) | 1.147*** | (0.184) | 3,715 | 0.043 |
| Pandemic/reopening midpoint | 0.031 | (0.252) | 1.073*** | (0.372) | 1,176 | 0.086 |
| 2022-onward midpoint | -0.774*** | (0.209) | 1.968*** | (0.215) | 1,658 | 0.120 |
| Midpoint outside 2021–2023 | -0.135 | (0.131) | 1.311*** | (0.153) | 4,599 | 0.055 |
| Point-forecast inflation (month-winsorized) | -0.142 | (0.166) | 1.862*** | (0.171) | 6,549 | 0.087 |
| Panel, survey-month fixed effects | -0.248*** | (0.084) | 1.235*** | (0.099) | 77,957 | 0.039 |

*Note:* Each row reports an interaction coefficient between standardized probability signed comovement and a standardized average inflation-expectation measure in the respondent-level expected-spending specifications of the main text, with the full control block; regressors are restandardized within each estimation sample. The inflation-level columns report the coefficient on the standardized average inflation expectation in the same regression and the implied inflation–spending slope at the mean of the interacted moderators in that estimation sample. The uncertainty rows add the respondent's average inflation-density interquartile range as a standardized control; the joint-specification rows include both the comovement-inflation and the uncertainty-inflation interactions in one regression and report each coefficient. The numeracy-interaction rows instead include the standardized numeracy share and its interaction with inflation expectations alongside the comovement interaction, replacing the raw numeracy-share control, and report each interaction coefficient. Timing rows split the sample by the median month-serial value of the respondent's retained measurement window, with cutoffs as in the timing-window diagnostic. The point-forecast row replaces the density-implied inflation mean with the respondent's average survey-month 1/99 winsorized signed point forecast. The panel row uses respondent-month observations of the trimmed spending outcome with survey-month fixed effects, the monthly inflation expectation as the interacted level, and standard errors clustered by respondent. Heteroskedasticity-robust (HC1) standard errors are in parentheses elsewhere. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.38 addresses household-sampling inference with a generated regressor. The comovement index is estimated from each respondent's 12 waves and then used as a regressor, so the analytic standard errors treat an estimated quantity as data. A respondent-level bootstrap re-draws households with replacement and re-runs the full estimation in every replication, including the index and inflation-level standardizations, the interaction term, the control imputation, and the OLS fit. Each drawn household enters with its own estimated 12-wave index; the bootstrap does not resample waves within household or re-estimate the index from alternative within-household histories. The bootstrap therefore quantifies cross-household sampling variability with the generated regressor carried through the specification, not a separate finite-panel measurement-error component. It confirms the analytic inference: the bootstrap standard errors (0.117 trimmed, 0.118 winsorized) are close to the HC1 values, the 95 percent percentile intervals exclude zero for both outcomes, and 0.3 percent or fewer of 1,000 replications produce a non-negative interaction. The bootstrap does not correct the attenuation from index noise, which biases the interaction toward zero, so the exercise is conservative in that limited sense.

Table C.38: Household bootstrap for the spending-slope interaction

| Outcome | Coefficient | HC1 SE | Bootstrap SE | 95% CI | Share ≥ 0 | *N* |
|---|---|---|---|---|---|---|
| Trimmed spending | -0.316 | 0.115 | 0.117 | [-0.531, -0.084] | 0.003 | 6,549 |
| Winsorized spending | -0.381 | 0.118 | 0.118 | [-0.607, -0.150] | 0.000 | 6,550 |

*Note:* Respondent-level (household) bootstrap for the comovement-inflation interaction in the spending-slope specification of Table 3.7. Each replication resamples respondents with replacement and re-runs the full estimation: the standardizations of the comovement index and of average inflation expectations, the interaction term, the mean-imputation of missing numeric controls, and the OLS fit are all recomputed within the resampled data. Each drawn household is included with its own estimated 12-wave index; the bootstrap does not resample waves within household or re-estimate the index from alternative within-household histories. It therefore quantifies cross-household sampling variability with the generated regressor carried through the specification, not a separate finite-panel measurement-error component, and it does not correct the attenuation from index noise, which biases the interaction toward zero. Monthly winsorization bounds for the winsorized outcome are held fixed at their full-sample values. The table reports the original point estimate with its analytic HC1 standard error, the bootstrap standard error, the 2.5th–97.5th percentile interval of the bootstrap distribution, and the share of replications with a non-negative interaction coefficient, based on 1,000 completed replications with a fixed seed.

Table C.39 reports a small secondary personal-risk block. The index does not meaningfully predict expected household-income growth, expected credit access, or debt-payment risk after the full controls and multiple-testing adjustment. It is negatively associated with job-loss expectations among respondents eligible for the job-loss question.

The job-loss association remains unexplained. Two candidate explanations do not account for it. First, the perceived-uncertainty channel predicts the wrong sign: mean job-loss probabilities are well below 50 (15.1 in this sample), so the uncertainty compression documented in Table C.25 would push job-loss answers up, not down, among high-comovement respondents. Second, a compositional story has no support in the observed controls. The question is routed only to employed respondents, and the association survives the demographic, education, income, expectation-level, and response-style controls, so it is not accounted for by the gradients in Table C.24. A selection explanation would have to operate through job characteristics the control set does not capture, such as occupation, industry, or tenure. I therefore treat the result as an unresolved secondary association rather than evidence for a mechanism.

Table C.39: Personal-risk prediction checks

| Outcome | Coef. | SE | p-value | q-value | $N$ | Mean |
|---|---|---|---|---|---|---|
| Expected household income growth | 0.031 | (0.103) | 0.764 | 0.764 | 6,544 | 3.809 |
| Expected credit harder | 0.006 | (0.008) | 0.468 | 0.624 | 6,550 | 3.200 |
| Debt-payment miss probability | -0.164 | (0.203) | 0.419 | 0.624 | 6,550 | 9.046 |
| Job-loss probability | -1.311*** | (0.279) | < 0.001 | < 0.001 | 4,307 | 15.120 |

*Note:* Each row regresses a respondent-level personal expectation on standardized probability signed comovement and the full control block. Income growth is the signed public field trimmed at absolute values above 50 percentage points. Expected credit access is recoded so higher values mean harder credit. Debt-payment and job-loss outcomes are probabilities on the 0–100 scale. Heteroskedasticity-robust (HC1) standard errors are in parentheses; q-values apply the Benjamini-Hochberg adjustment across the four personal-risk rows. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

# Appendix D: Experiment Instructions

**Summary**

Section D.1 provides the instructions for the human validation study.
Section D.2 provides the instructions for the demand experiment.
Section D.3 provides the instructions for the usage incentive experiment.

## D.1 Human Validation Study

### D.1.1 Wave 1

This study is conducted by researchers from the Frankfurt School of Finance & Management, NHH Norwegian School of Economics, University of Cologne, and MIT. You must be at least 18 years of age to participate in this study. If you do not fulfill these requirements, please do not continue any further.

The goal of this study is to assess newspaper articles. We do not anticipate any risks or discomforts from participating in this study.

Your responses to this survey will be anonymous; we will not collect your name or any other personal identifiers. We will only collect your Prolific ID to track your participation in this study, administer potential bonus payments, and invite you to potential future follow-up studies.

You are not allowed to participate in this study more than once. If you experience a technical error or problem, do not try to restart or retake the study. Rather, send us an email with a description of your problem and we will get back to you.

If you have any questions regarding this study, please email roth@wiso.uni-koeln.de

If you feel you have been treated unfairly, or you have questions regarding your rights as a research subject, you may contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, M.I.T., Room E25-143b, 77 Massachusetts Ave, Cambridge, MA 02139, phone 1-617-253-6787.

Do you consent to participate in this study?

- ○ I consent
- ○ I do not consent and wish to exit the study

→

This question is about the following problem. Sometimes there are participants who do not carefully read the questions and just quickly click through the survey. This means that there are a lot of random answers which compromise the results of research studies. To show that you read our questions carefully, please choose both "Very strongly interested" and "Not at all interested" as your answer to the question below.

How interested are you in sports?

- [ ] Very strongly interested
- [ ] Somewhat interested
- [ ] A little interested
- [ ] Almost not interested
- [ ] Not at all interested

→

**Instructions**

You will see pairs of newspaper articles. For each pair, you will be asked to compare them along one of the following dimensions:

- **Entertaining:** use of humor, puns, clever word choices, and a more playful narrative style
- **Liberal perspective:** use of language that is more commonly used by liberals and highlights issues that matter to liberals
- **Conservative perspective:** use of language that is more commonly used by conservatives and highlights issues that matter to conservatives
- **Positive:** use of more positive language and highlighting positive aspects
- **Negative:** use of more negative language and highlighting negative aspects
- **Less complex:** use of simpler terms and brief explanations for unfamiliar concepts

We provide examples of the types of questions that you will encounter below.

**Example: conservative perspective**

| Article A | Article B |
|---|---|
| **Trump Targets Black Voters and Controversial Conservative Group in Michigan Campaign Push**<br><br>Donald Trump plans to court both Black voters and a conservative group with alleged ties to white supremacists in a strategic Michigan visit to boost his 2024 election bid. | **Trump Rallies Black Voters and Conservatives in Key Michigan Stops**<br><br>Donald Trump aims to unite Black voters and conservatives with strong economic and border security messages during key events in Michigan, challenging Biden in a crucial state. |

Which article is more aligned with a conservative perspective?

- ○ Article A is more aligned with a conservative perspective
- ○ Article B is more aligned with a conservative perspective

**Example: entertaining**

| Article A | Article B |
|---|---|
| **Trump's Michigan Gambit: Wooing Black Voters and Wading into Extremist Waters**<br><br>Trump sets out on a bold Michigan tour to woo Black voters and speak at a controversial conservative convention, highlighting his strategy to dent Biden's diverse voter base in this battleground state. | **Trump Targets Black Voters and Controversial Conservative Group in Michigan Campaign Push**<br><br>Donald Trump plans to court both Black voters and a conservative group with alleged ties to white supremacists in a strategic Michigan visit to boost his 2024 election bid. |

- ○ Article A is more entertaining
- ○ Article B is more entertaining

**Example: positive**

| Article A | Article B |
|---|---|
| **Trump Targets Black Voters and Controversial Conservative Group in Michigan Campaign Push**<br><br>Donald Trump plans to court both Black voters and a conservative group with alleged ties to white supremacists in a strategic Michigan visit to boost his 2024 election bid. | **Trump Engages Black Voters and Conservative Leaders in Michigan Push**<br><br>Donald Trump plans to engage Black voters and connect with conservative leaders in Michigan, highlighting his strategy to unite diverse voter groups and challenge President Biden in a key state. |

Which article is more positive?

- ◯ Article A is more positive
- ◯ Article B is more positive

**Example: less complex**

| Article A | Article B |
|---|---|
| **Trump Targets Black Voters and Controversial Conservative Group in Michigan Campaign Push**<br><br>Donald Trump plans to court both Black voters and a conservative group with alleged ties to white supremacists in a strategic Michigan visit to boost his 2024 election bid. | **Trump Courts Black Voters and Conservatives in Michigan**<br><br>Donald Trump plans to speak at a Black church in Detroit and at a conservative convention tied to extremists, aiming to unite different voter groups ahead of the next election. |

Which article is less complex?

- ◯ Article A is less complex
- ◯ Article B is less complex

**Example: liberal perspective**

| Article A | Article B |
|---|---|
| **Trump's Michigan Tour Targets Black Voters While Embracing Controversial Conservative Group**<br><br>Donald Trump aims to sway Black voters with economic and border security messages in a crucial Michigan visit, while aligning with Turning Point Action, a group criticized for extremist ties, to consolidate conservative support. | **Trump Targets Black Voters and Controversial Conservative Group in Michigan Campaign Push**<br><br>Donald Trump plans to court both Black voters and a conservative group with alleged ties to white supremacists in a strategic Michigan visit to boost his 2024 election bid. |

Which article is more aligned with a liberal perspective?

- ○ Article A is more aligned with a liberal perspective
- ○ Article B is more aligned with a liberal perspective

**Example: negative**

| Article A | Article B |
|---|---|
| **Trump Courts Black Voters While Embracing Extremists**<br><br>In a desperate bid to win over Black voters, Donald Trump is simultaneously aligning with a controversial conservative group linked to white supremacists during his Michigan stops, highlighting his struggle to unite disparate voter groups against Joe Biden. | **Trump Targets Black Voters and Controversial Conservative Group in Michigan Campaign Push**<br><br>Donald Trump plans to court both Black voters and a conservative group with alleged ties to white supremacists in a strategic Michigan visit to boost his 2024 election bid. |

Which article is more negative?

- ○ Article A is more negative
- ○ Article B is more negative

→

On the next few pages, you will be asked to rate **6 article pairs** along **one dimension for each**. It is important that you take this task seriously.

→

**Round 1/6: Which article is more aligned with a conservative perspective?**

| Article A | Article B |
|---|---|
| **Judge Orders DOJ to Uncover Pence Probe Details in Trump Election Case**<br><br>*In a significant development in Donald Trump's election interference case, a federal judge has directed prosecutors to search for Justice Department information related to Mike Pence's handling of classified documents, which Trump's defense claims could be pivotal.*<br><br>In the election interference case against Donald Trump, a federal judge has instructed prosecutors to search for and provide any Justice Department information related to a separate investigation into Mike Pence's handling of classified documents. This decision follows an argument from Trump's lawyers that such information could be crucial to their defense, particularly if it reveals that Pence had motives to implicate Trump while facing his own investigation.<br>Although special counsel Jack Smith's team claims no involvement in the Pence investigation beyond public knowledge, U.S. Dis...<br><br>Show more | **Judge Orders DOJ to Disclose Evidence in Trump-Pence Investigation**<br><br>*A federal judge has directed prosecutors to reveal Justice Department information on Mike Pence's classified documents probe, supporting Trump's defense claims of potential bias in the politically charged election interference case.*<br><br>In the election interference case against Donald Trump, a federal judge has instructed prosecutors to search for and provide any Justice Department information related to a separate investigation into Mike Pence's handling of classified documents. This decision follows an argument from Trump's legal team that such information could be essential for a fair defense, especially if it uncovers any motives Pence might have had to unduly implicate Trump while under his own investigation.<br>Although special counsel Jack Smith's team claims no involvement in the Pence investigation beyond publ...<br><br>Show more |

Which article is more aligned with a conservative perspective?

- ◯ Article A is more conservative
- ◯ Article B is more conservative

→

**Round 2/6: Which article is more aligned with a conservative perspective?**

| Article A | Article B |
|---|---|
| **Georgia's Record Early Voting Reflects Commitment to Election Integrity**<br><br>*Georgia's early voting surge, with over 540,000 ballots cast, highlights the state's dedication to securing election integrity with stringent mail-in voting measures, as voters express strong partisan motivations.*<br><br>In Georgia, early voting has kicked off with remarkable enthusiasm, as over 540,000 ballots have been cast by Wednesday, surpassing 10% of the anticipated voter turnout for this election cycle. This early surge includes over 310,000 people voting in person on Tuesday, setting a new record for first-day early voting turnout, and an additional 180,000 by Wednesday afternoon. Additionally, 30,000 mail-in ballots have been accepted. This follows the significant participation of 5 million Georgians in the 2020 presidential election, underscoring the vital role Georgia plays as a battleground sta...<br><br>Show more | **Georgia's Early Voting Sets Record Pace as Voters Navigate New Rules and Political Divides**<br><br>*Over 540,000 ballots have already been cast in Georgia's early voting, highlighting intense voter interest amid new voting regulations and deep partisan divides in the battleground state.*<br><br>In Georgia, early voting is off to a strong start with over 540,000 ballots cast by Wednesday, surpassing 10% of the expected voter turnout for this election cycle. This early surge includes more than 310,000 people voting in person on Tuesday, marking the highest first-day early voting turnout on record, and another 180,000 by Wednesday afternoon. Additionally, 30,000 mail-in ballots have been accepted. This comes in the context of the 5 million Georgians who voted in the 2020 presidential election, highlighting the intensity and importance of this electoral cycle in the battleground state...<br><br>Show more |

Which article is more aligned with a conservative perspective?

- ○ Article A is more conservative
- ○ Article B is more conservative

→

**Round 3/6: Which article is less complex?**

| Article A | Article B |
|---|---|
| **Protesters Arrested at New York Rally Against Israeli Actions**<br><br>*About 200 people were arrested during a protest outside the New York Stock Exchange against Israel's military actions in Gaza, organized by Jewish Voice for Peace, calling for an end to U.S. support of Israel and highlighting tensions in the ongoing conflict.*<br><br>About 200 people were arrested while protesting against Israel's military actions in Gaza outside the New York Stock Exchange. The protest was organized by a group called Jewish Voice for Peace. Protesters chanted slogans like "Let Gaza live!" to urge the U.S. government to stop supporting Israel's military actions. The protesters accused the U.S. of helping what they call a genocide of Palestinians. They said the U.S. sends bombs that are used against people in Gaza, and that American weapons companies are making money from this.<br>The protesters did not enter the stock exchange build...<br><br>Show more | **Protesters Arrested at New York Stock Exchange Rally Against U.S. Support for Israel**<br><br>*Approximately 200 demonstrators were arrested during a protest organized by Jewish Voice for Peace outside the New York Stock Exchange, as they called for an end to U.S. support for Israel's military actions in Gaza amid escalating tensions in the region.*<br><br>Around 200 demonstrators were arrested during a protest against Israel's military actions in Gaza outside the New York Stock Exchange. The protest, organized by Jewish Voice for Peace, featured chants such as "Let Gaza live!" and focused on urging the U.S. government to cease supporting Israel's military efforts. The demonstrators accused the U.S. of contributing to what they termed a genocide of Palestinians by supplying bombs, which they claim are being used to kill communities in Gaza, while U.S. weapons manufacturers benefit financially.<br>Although none of the protesters breached t...<br><br>Show more |

Which article is less complex?

- ○ Article A is less complex
- ○ Article B is less complex

→

**Round 4/6: Which article is more aligned with a liberal perspective?**

| Article A | Article B |
|---|---|
| **Trump Targets California Base with Strategic Coachella Visit Ahead of Election Day**<br><br>*Despite California's Democratic leanings, Donald Trump's visit to Coachella aims to energize his supporter base, raise funds, and bolster Republican efforts in key nationwide races.*<br><br>Donald Trump's decision to visit California, a solidly Democratic stronghold, just weeks before Election Day seems puzzling given his slim chances of winning the state. However, there are strategic reasons for his stop in Coachella, a city known for its famous music festival. Despite losing California by a wide margin in 2020, Trump garnered over six million votes, with strong support in rural counties. This large base of supporters serves as potential volunteers for state races and phone banking efforts targeting more contested states.<br>The visit also ensures extensive media coverage...<br><br>Show more | **Trump Courts California Conservatives as Election Day Nears Lead: Despite facing an uphill battle in California, Trump's visit to Coachella aims to galvanize his base, boost fundraising, and draw media attention, highlighting contrasts with Democratic leadership while energizing Republican supporters.**<br><br>*Lead: Despite facing an uphill battle in California, Trump's visit to Coachella aims to galvanize his base, boost fundraising, and draw media attention, highlighting contrasts with Democratic leadership while energizing Republican supporters.*<br><br>Donald Trump's decision to visit California, a bastion of Democratic values, just weeks before Election Day raises questions, especially considering his minimal chances of success in the state. However, there are strategic reasons for his stop in Coachella, a city known for its famous music festival. Despite losing California by a wide margin in 2020, Trump garnered over six million votes, with strong support in rural counties. This large base of supporters serves as potential volunteers for state races and phone banking efforts targeting more contested states. The visit also ensures extens...<br><br>Show more |

Which article is more aligned with a liberal perspective?

- ○ Article A is more liberal
- ○ Article B is more liberal

→

**Round 5/6: Which article is more aligned with a conservative perspective?**

| Article A | Article B |
|---|---|
| **Kamala Harris' Campaign Faces Gender Bias Hurdle: Engaging Male Voters Remains a Challenge**<br><br>*As Kamala Harris' presidential campaign navigates potential gender biases, her team intensifies efforts to engage male voters amid concerns over apathy and misinformation.*<br><br>Concerns have emerged about potential gender bias affecting Kamala Harris' candidacy as she seeks to break barriers as a female presidential candidate. While Harris seldom discusses her gender on the campaign trail, her supporters are actively appealing to male voters to combat sexism and apathy. Former President Barack Obama has addressed this issue, particularly targeting Black men, suggesting that reluctance may stem from discomfort with the idea of a woman in the presidency. Actor Ed O'Neill also calls on men to support Harris by voting for a woman. Efforts to engage Hispanic mal...<br><br>Show more | **Gender Politics Overshadow Policy in Kamala Harris' Presidential Campaign**<br><br>*Kamala Harris' campaign is criticized for focusing on gender bias issues, overshadowing critical policy discussions, as attempts to sway male voters through identity politics distract from her qualifications and record.*<br><br>Concerns are being raised about potential gender bias influencing Kamala Harris' candidacy as she attempts to make history as a female presidential candidate, a focus that diverts from addressing core policy issues. While Harris seldom discusses her gender on the campaign trail, her supporters are actively appealing to male voters to combat sexism and apathy.<br>Former President Barack Obama has weighed in on this topic, focusing on Black men, suggesting that any reluctance could be tied to unease with a female president, rather than considering valid policy concerns. Actor Ed O'Neill u...<br><br>Show more |

Which article is more aligned with a conservative perspective?

- ○ Article A is more conservative
- ○ Article B is more conservative

→

**Round 6/6: Which article is more entertaining?**

| Article A | Article B |
|---|---|
| **Election 2024: Trump and Harris Court Male Voters with Diverging Visions of Masculinity**<br><br>*As the November 5th election approaches, Donald Trump's and Kamala Harris's campaigns are intensifying efforts to win over male voters by presenting contrasting narratives of masculinity and identity politics.*<br><br>In the lead-up to the November 5th election, Donald Trump and Kamala Harris's campaigns are heavily targeting male voters, highlighting a significant gender and identity politics dynamic. Trump's strategy employs a hypermasculine tone, appealing to traditional gender roles and casting a vote for him as a symbol of American manliness. This approach is evident in his appearances on podcasts and gaming platforms, where figures like Charlie Kirk assert that voting for Trump is a measure of masculinity. Trump's narrative positions him as a protector, particularly appealing to suburban women whil...<br><br>Show more | **Battle of the Genders: Trump vs. Harris in Macho Showdown**<br><br>*As Trump amps up the macho appeal and Harris courts diverse men with relatable charm, the fierce battle for male votes unfolds in a gender-divided election drama.*<br><br>As the countdown to the November 5th election intensifies, Donald Trump and Kamala Harris are locked in a fierce battle for the hearts—and votes—of the nation's male populace. Trump is turning up the macho factor to eleven, crafting a narrative where casting a vote for him is akin to striking a heroic pose on the cover of an all-American comic book. This testosterone-fueled rally cry echoes across podcasts and gaming platforms, with cheerleaders like Charlie Kirk declaring that a vote for Trump is the ultimate test of manhood. Trump paints himself as the gallant protector, hoping to win ove...<br><br>Show more |

Which article is more entertaining?

- ◯ Article A is more entertaining
- ◯ Article B is more entertaining

→

How do you primarily consume news?

- ○ Online
- ○ TV
- ○ Radio
- ○ Print news

Which of the newspapers below have you read at least once during the last 12 months? Please select all that apply.

- ☐ Breitbart
- ☐ BuzzFeed News
- ☐ Boston Herald
- ☐ Chicago Sun-Times
- ☐ Daily Mail
- ☐ Drudge Report
- ☐ InfoWars
- ☐ Los Angeles Times
- ☐ New Republic
- ☐ Newsmax
- ☐ New York Daily News
- ☐ New York Post
- ☐ Palmer Report
- ☐ The Denver Post
- ☐ The Huffington Post
- ☐ The Mercury News
- ☐ The New York Times
- ☐ The Wall Street Journal
- ☐ The Washington Post
- ☐ The Washington Times
- ☐ USA Today
- ☐ I have not read any of the newspapers above during the past 12 months

How often do you watch or read politics and economic news?

- Frequently
- Somewhat frequently
- Occasionally
- Rarely
- Never

How interested are you in political and economic news?

- Very interested
- Interested
- Somewhat interested
- Not interested
- Not at all interested

How much do you trust news in general?

- A great deal
- A lot
- A moderate amount
- A little
- None at all

→

What is your gender?

- ○ Male
- ○ Female

What is your age?

Which category best describes your highest level of education?

- ○ Did not graduate high school
- ○ High school degree/GED
- ○ Some college
- ○ 2-year college degree
- ○ 4-year college degree
- ○ Master's degree
- ○ Doctoral degree
- ○ Professional degree (JD, MD, MBA)

Which of the following best describes your race or ethnicity?

- ○ African American/Black
- ○ Asian/Asian American
- ○ Caucasian/White
- ○ Native American, Inuit or Aleut
- ○ Native Hawaiian/Pacific Islander
- ○ Other

Are you of Hispanic, Latino, or Spanish origin?

- ○ Yes
- ○ No

What was your household's gross income in 2023?

- Less than $15,000
- $15,000 to $24,999
- $25,000 to $49,999
- $50,000 to $74,999
- $75,000 to $99,999
- $100,000 to $149,999
- $150,000 to $199,999
- $200,000 or more

In politics, as of today, do you consider yourself a Republican, a Democrat, or an Independent?

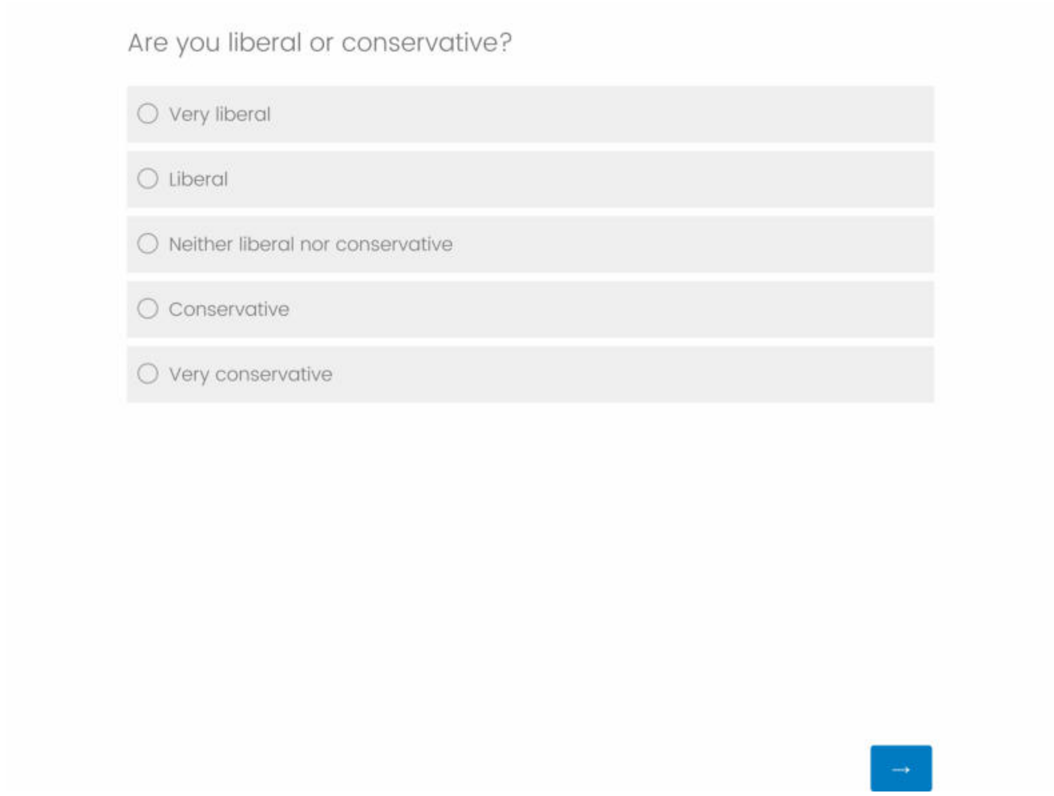
- Republican
- Democrat
- Independent

What is your region of residence?

- Northeast (CT, ME, MA, NH, RI, VT, NJ, NY, PA)
- Midwest (IL, IN, MI, OH, WI, IA, KS, MK, MO, NE, ND, SD)
- South (DE, DC, FL, GA, MD, NC, SC, VA, WV, AL, KY, MS, TN, AR, LA, OK, TX)
- West (AZ, CO, ID, NM, MT, UT, NV, WY, AK, CA, HI, OR, WA)

What is your current employment status?

- Full-time employee
- Part-time employee
- Self-employed or small business owner
- Unemployed and looking for work
- Student
- Not in labor force (for example: retired or full-time parent)

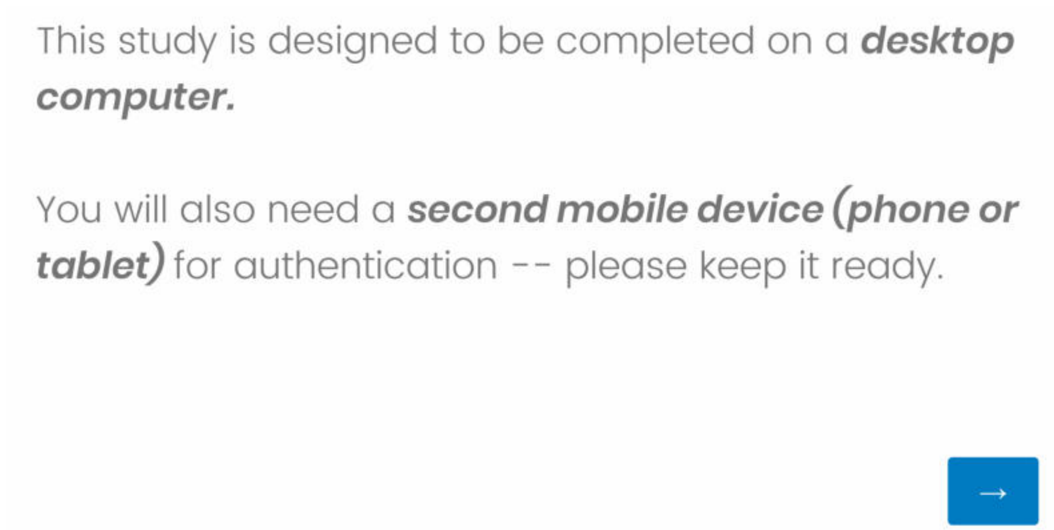


Are you liberal or conservative?

- Very liberal
- Liberal
- Neither liberal nor conservative
- Conservative
- Very conservative

→

### D.1.2 Wave 2



This study is designed to be completed on a ***desktop computer.***

You will also need a ***second mobile device (phone or tablet)*** for authentication -- please keep it ready.

→

## Please confirm

**Response quality is important for the quality of this research project. For this reason, we will review responses and reject respondents who**

- speed through the survey
- submit invalid data
- use bots or artificial intelligence (AI, e.g. ChatGPT or Google Gemini)

You don't need these AI tools anyway. This survey is all about your own views and thoughts.

**Please respond carefully to this survey.**

○ I confirm

→

## Authentication

**To help us prevent bots and automated programs from participating in this study, we ask you to verify your participation using a second device.**

**Here's how it works:**
On the next page, a QR code will appear. Please scan it with a second device (such as a smartphone or tablet). This will open a secure webpage on that device, confirming your participation. Because scanning a QR code requires human action, this helps us verify that you are a real person.

You can complete this step only once, and it must be done within 60 seconds. It is required in order to continue with the survey.

→


Scan this code with your second device to continue ...
Waiting for phone …
0057

This study is conducted by researchers from the Frankfurt School of Finance & Management, NHH Norwegian School of Economics, University of Cologne, and MIT. You must be at least 18 years of age to participate in this study. If you do not fulfill these requirements, please do not continue any further.

The goal of this study is to assess newspaper articles. We do not anticipate any risks or discomforts from participating in this study.

Your responses to this survey will be anonymous; we will not collect your name or any other personal identifiers. We will only collect your Prolific ID to track your participation in this study, administer potential bonus payments, and invite you to potential future follow-up studies.

You are not allowed to participate in this study more than once. If you experience a technical error or problem, do not try to restart or retake the study. Rather, send us an email with a description of your problem and we will get back to you.

If you have any questions regarding this study, please email roth@wiso.uni-koeln.de

If you feel you have been treated unfairly, or you have questions regarding your rights as a research subject, you may contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, M.I.T., Room E25-143b, 77 Massachusetts Ave, Cambridge, MA 02139, phone 1-617-253-6787.

Do you consent to participate in this study?

- ○ I consent
- ○ I do not consent and wish to exit the study

→

This question is about the following problem. Sometimes there are participants who do not carefully read the questions and just quickly click through the survey. This means that there are a lot of random answers which compromise the results of research studies. To show that you read our questions carefully, please choose both "Very strongly interested" and "Not at all interested" as your answer to the question below.

How interested are you in sports?

- [ ] Very strongly interested
- [ ] Somewhat interested
- [ ] A little interested
- [ ] Almost not interested
- [ ] Not at all interested

→

## Instructions

You will see pairs of newspaper articles. For each pair, you will be asked to compare them along one of the following dimensions:

- **Entertaining:** use of humor, puns, clever word choices, and a more playful narrative style
- **Liberal perspective:** use of language that is more commonly used by liberals and highlights issues that matter to liberals
- **Conservative perspective:** use of language that is more commonly used by conservatives and highlights issues that matter to conservatives
- **Positive:** use of more positive language and highlighting positive aspects
- **Negative:** use of more negative language and highlighting negative aspects
- **Less complex:** use of simpler terms and brief explanations for unfamiliar concepts
- **Facts only:** use of fact-driven language and removing of opinions

We provide examples of the types of questions that you will encounter below.

## Example: conservative perspective

| Article A | Article B |
|---|---|
| **Trump Targets Black Voters and Controversial Conservative Group in Michigan Campaign Push**<br><br>Donald Trump plans to court both Black voters and a conservative group with alleged ties to white supremacists in a strategic Michigan visit to boost his 2024 election bid. | **Trump Rallies Black Voters and Conservatives in Key Michigan Stops**<br><br>Donald Trump aims to unite Black voters and conservatives with strong economic and border security messages during key events in Michigan, challenging Biden in a crucial state. |

Which article is more aligned with a conservative perspective?

- ○ Article A is more aligned with a conservative perspective
- ○ Article B is more aligned with a conservative perspective

**Example: entertaining**

| Article A | Article B |
|---|---|
| **Trump's Michigan Gambit: Wooing Black Voters and Wading into Extremist Waters**<br><br>Trump sets out on a bold Michigan tour to woo Black voters and speak at a controversial conservative convention, highlighting his strategy to dent Biden's diverse voter base in this battleground state. | **Trump Targets Black Voters and Controversial Conservative Group in Michigan Campaign Push**<br><br>Donald Trump plans to court both Black voters and a conservative group with alleged ties to white supremacists in a strategic Michigan visit to boost his 2024 election bid. |

- ◯ Article A is more entertaining
- ◯ Article B is more entertaining

**Example: negative**

| Article A | Article B |
|---|---|
| **Trump Courts Black Voters While Embracing Extremists**<br><br>In a desperate bid to win over Black voters, Donald Trump is simultaneously aligning with a controversial conservative group linked to white supremacists during his Michigan stops, highlighting his struggle to unite disparate voter groups against Joe Biden. | **Trump Targets Black Voters and Controversial Conservative Group in Michigan Campaign Push**<br><br>Donald Trump plans to court both Black voters and a conservative group with alleged ties to white supremacists in a strategic Michigan visit to boost his 2024 election bid. |

Which article is more negative?

- ◯ Article A is more negative
- ◯ Article B is more negative

**Example: less complex**

| Article A | Article B |
|---|---|
| **Trump Targets Black Voters and Controversial Conservative Group in Michigan Campaign Push**<br><br>Donald Trump plans to court both Black voters and a conservative group with alleged ties to white supremacists in a strategic Michigan visit to boost his 2024 election bid. | **Trump Courts Black Voters and Conservatives in Michigan**<br><br>Donald Trump plans to speak at a Black church in Detroit and at a conservative convention tied to extremists, aiming to unite different voter groups ahead of the next election. |

Which article is less complex?

- ◯ Article A is less complex
- ◯ Article B is less complex

**Example: facts only**

| Article A | Article B |
|---|---|
| **Trump Targets Black Voters and Controversial Conservative Group in Michigan Campaign Push**<br><br>Donald Trump plans to court both Black voters and a conservative group with alleged ties to white supremacists in a strategic Michigan visit to boost his 2024 election bid. | **Trump Meets Black Voters and Conservative Group in Michigan Campaign Stop**<br><br>Donald Trump will meet with Black voters and a conservative group in Michigan during a campaign visit for the 2024 presidential election. The group has been accused of having ties to white supremacists. |

Which article is facts only?

- ◯ Article A is facts only
- ◯ Article B is facts only

**Example: liberal perspective**

| Article A | Article B |
|---|---|
| **Trump's Michigan Tour Targets Black Voters While Embracing Controversial Conservative Group**<br><br>Donald Trump aims to sway Black voters with economic and border security messages in a crucial Michigan visit, while aligning with Turning Point Action, a group criticized for extremist ties, to consolidate conservative support. | **Trump Targets Black Voters and Controversial Conservative Group in Michigan Campaign Push**<br><br>Donald Trump plans to court both Black voters and a conservative group with alleged ties to white supremacists in a strategic Michigan visit to boost his 2024 election bid. |

Which article is more aligned with a liberal perspective?

- ◯ Article A is more aligned with a liberal perspective
- ◯ Article B is more aligned with a liberal perspective

**Example: positive**

| Article A | Article B |
|---|---|
| **Trump Targets Black Voters and Controversial Conservative Group in Michigan Campaign Push**<br><br>Donald Trump plans to court both Black voters and a conservative group with alleged ties to white supremacists in a strategic Michigan visit to boost his 2024 election bid. | **Trump Engages Black Voters and Conservative Leaders in Michigan Push**<br><br>Donald Trump plans to engage Black voters and connect with conservative leaders in Michigan, highlighting his strategy to unite diverse voter groups and challenge President Biden in a key state. |

Which article is more positive?

- ◯ Article A is more positive
- ◯ Article B is more positive

→

On the next few pages, you will be asked to rate **6 article pairs** along **one dimension for each**. It is important that you take this task seriously.

→

**Round 1/6: Which article is more positive?**

| Article A | Article B |
|---|---|
| **Tennessee Expands School Choice with Transformative Voucher Program**<br><br>*Tennessee's expanded school voucher program will provide 20,000 families with increased educational options and enhanced parental rights, marking a historic step in school choice.*<br><br>Tennessee lawmakers have approved a transformative expansion of the state's school voucher program, offering more families access to public funds for private school tuition, regardless of income. The legislation, backed enthusiastically by Republican Governor Bill Lee and supported by President Trump's advocacy, will provide 20,000 vouchers of approximately $7,000 each starting next year. Half of these vouchers will be reserved for lower-income families, students with disabilities, and other eligible participants, while the remaining vouchers will be available to any student eligible to att...<br><br>Show more | **Tennessee Expands School Voucher Program, Opens Access to All Students**<br><br>*Tennessee has approved a major expansion of its school voucher program, granting public funds for private education to all eligible families, sparking heated debate over its impact on public schools.*<br><br>Tennessee lawmakers approved a significant expansion of the state's school voucher program, allowing more families to use public funds for private school tuition regardless of income. The legislation, strongly supported by Republican Governor Bill Lee and bolstered by President Trump's advocacy, includes 20,000 vouchers of approximately $7,000 each starting next year. Half will be reserved for lower-income families, students with disabilities, or other eligible participants, while the remaining vouchers will be available to any student eligible to attend public schools.<br>The approval ...<br><br>Show more |

Which article is more positive?

- ○ Article A is more positive
- ○ Article B is more positive

→

**Round 2/6: Which article is facts only?**

| Article A | Article B |
|---|---|
| **President Trump to Use Guantanamo Bay for Migrant Detention Amid Capacity Strains**<br><br>*President Trump plans to detain tens of thousands of migrants at Guantanamo Bay as part of his administration's ongoing efforts to address immigration challenges and capacity limits.*<br><br>President Donald Trump announced plans to use the detention center at Guantanamo Bay, Cuba, to hold tens of thousands of migrants who cannot be returned to their home countries. This decision aligns with Trump's continued emphasis on deportation as a key policy priority.<br>The naval base at Guantanamo Bay, best known for housing terrorism suspects post-9/11, also contains the Migrant Operations Center, which has been used for decades to detain migrants intercepted at sea, primarily from Haiti and Cuba. Reports from advocacy groups have criticized the conditions at the facility, describ...<br><br>Show more | **President Trump Plans to Use Guantanamo for Migrant Detention**<br><br>*President Trump announced plans to use Guantanamo Bay's facilities to detain tens of thousands of migrants who cannot be deported to their home countries.*<br><br>President Donald Trump announced plans to use the detention center at Guantanamo Bay, Cuba, to hold tens of thousands of migrants who cannot be returned to their home countries. This decision reflects the administration's focus on deportation as a policy priority.<br>The naval base at Guantanamo Bay, best known for housing terrorism suspects post-9/11, also contains the Migrant Operations Center, which has been used for decades to detain migrants intercepted at sea, primarily from Haiti and Cuba. Advocacy groups have reported concerns about the conditions at the facility, describing the...<br><br>Show more |

Which article is more fact-driven?

- ○ Article A is more fact-driven
- ○ Article B is more fact-driven

→

**Round 3/6: Which article is more negative?**

| Article A | Article B |
|---|---|
| **Netanyahu Seeks Meeting with Trump Amid Middle East Tensions** | **Netanyahu May Meet Trump in Washington Amid Ceasefire Diplomacy** |
| *Israeli Prime Minister Benjamin Netanyahu is making efforts to secure a meeting with President Donald Trump in Washington as early as next week, despite the uncertainty surrounding the plans.* | *Israeli Prime Minister Benjamin Netanyahu is considering a meeting with President Donald Trump at the White House next week, as efforts to reinforce Middle East ceasefires continue.* |
| Israeli Prime Minister Benjamin Netanyahu is desperately attempting to arrange a meeting with President Donald Trump in Washington as early as next week, according to U.S. officials involved in preliminary trip planning. If finalized, Netanyahu could gain the symbolic but largely superficial distinction of being the first foreign leader to meet Trump at the White House since his inauguration. The planning remains highly uncertain, and details may solidify during a visit to Israel by Trump's special Middle East envoy, Steve Witkoff, this week. The White House has yet to comment, and N... | Israeli Prime Minister Benjamin Netanyahu is exploring the possibility of meeting President Donald Trump in Washington as early as next week, according to U.S. officials involved in preliminary trip planning. If finalized, Netanyahu would become the first foreign leader to meet Trump at the White House since his inauguration. The planning remains uncertain, and details may solidify during a visit to Israel by Trump's special Middle East envoy, Steve Witkoff, this week. The White House has yet to comment, and Netanyahu's spokesperson, Omer Dostri, confirmed no official invitation has ... |
| Show more | Show more |

Which article is more negative?

- ◯ Article A is more negative
- ◯ Article B is more negative

→

**Round 4/6: Which article is facts only?**

| Article A | Article B |
|---|---|
| **Judge Temporarily Halts Trump Administration Federal Funding Freeze**<br><br>*A federal judge temporarily blocked the Trump administration's efforts to freeze federal funding, mandating uninterrupted allocation to states under existing laws and regulations.*<br><br>A federal judge has issued a temporary hold on the Trump administration's efforts to freeze federal funding, related to the allocation of trillions of dollars in grants and loans. Judge John McConnell ruled in favor of nearly two dozen states seeking to prevent federal agencies from halting funding while the court considers their request for a preliminary injunction. His order mandates that federal financial assistance to the states must continue uninterrupted, restricted only by existing laws, regulations, and terms.<br>This ruling follows the Office of Management and Budget's withdraw...<br><br>Show more | **Federal Judge Blocks Trump Administration's Attempt to Freeze Funding**<br><br>*A federal judge has temporarily halted the Trump administration's effort to freeze federal funding, ensuring continued financial assistance to states while legal challenges proceed.*<br><br>A federal judge has issued a temporary hold on the Trump administration's efforts to freeze federal funding, marking another development in the battle over the allocation of trillions of dollars in grants and loans. Judge John McConnell ruled in favor of nearly two dozen states seeking to prevent federal agencies from halting funding while the court considers their request for a preliminary injunction. His order mandates that federal financial assistance to the states must continue uninterrupted, restricted only by existing laws, regulations, and terms.<br>This ruling comes even after t...<br><br>Show more |

Which article is more fact-driven?

- ○ Article A is more fact-driven
- ○ Article B is more fact-driven

→

**Round 5/6: Which article is more negative?**

| Article A | Article B |
|---|---|
| **Trump Deflects Blame with Baseless Diversity Hiring Claims After Crash**<br><br>*President Trump irresponsibly linked diversity hiring to an aviation crash without evidence, deflecting blame and offering no clear plan to enhance safety.*<br><br>President Trump began a White House briefing with a perfunctory moment of silence for the victims of a collision between an American Airlines plane and a U.S. Army helicopter at Reagan National Airport, before quickly turning the focus away from the tragedy. However, his remarks quickly derailed into unfounded accusations against diversity hiring, irresponsibly insinuating that such policies might have contributed to the crash without providing any evidence. He cited previous administration efforts to diversify the Federal Aviation Administration (FAA) workforce, claiming they compromised s...<br><br>Show more | **President Trump Links FAA Diversity Policies to Reagan Airport Crash Without Evidence**<br><br>*President Trump criticized past diversity policies during a briefing on a deadly aircraft collision at Reagan National Airport, despite authorities not yet determining the cause of the crash.*<br><br>President Trump began a White House briefing with a moment of silence for the victims of a collision between an American Airlines plane and a U.S. Army helicopter at Reagan National Airport. However, his remarks soon shifted to criticizing diversity hiring and suggesting, without evidence, that such policies might have contributed to the crash. He cited previous administration efforts to diversify the Federal Aviation Administration (FAA) workforce, claiming they compromised standards, though he provided no supporting details. Authorities are still investigating the cause of the inci...<br><br>Show more |

Which article is more negative?

- ○ Article A is more negative
- ○ Article B is more negative

→

**Round 6/6: Which article is more aligned with a liberal perspective?**

| Article A | Article B |
|---|---|
| **Trump's Baseless Claims on FAA Diversity Spark Backlash**<br><br>*President Trump has faced criticism for linking an aviation tragedy to FAA diversity policies without evidence, reigniting debates over inclusion in federal agencies.*<br><br>President Donald Trump has controversially linked the recent aviation disaster—the deadliest in over twenty years—to a Federal Aviation Administration (FAA) diversity hiring program, a claim that has been met with widespread criticism for its lack of evidence and divisive rhetoric. The collision between an American Airlines regional jet and an Army Black Hawk helicopter on Wednesday claimed 67 lives. Despite Trump's claims, no evidence has emerged connecting the FAA's diversity efforts to the crash. These comments have reignited debates around diversity policies and their role in federal ag...<br><br>Show more | **President Trump Links FAA Diversity Program to Deadly Aviation Crash**<br><br>*President Trump has attributed the recent aviation tragedy to FAA diversity hiring policies, sparking controversy despite a lack of evidence supporting the claim.*<br><br>President Donald Trump has linked the recent aviation disaster—the deadliest in over twenty years—to a Federal Aviation Administration (FAA) diversity hiring program, suggesting it compromised safety. The collision between an American Airlines regional jet and an Army Black Hawk helicopter on Wednesday claimed 67 lives. Despite Trump's claims, no evidence has emerged connecting the FAA's diversity efforts to the crash. These comments have reignited debates around diversity policies and their role in federal agencies. The FAA has long faced a shortage of air traffic controllers, an is...<br><br>Show more |

Which article is more aligned with a liberal perspective?

- ◯ Article A is more liberal
- ◯ Article B is more liberal

→

How do you primarily consume news?

- ○ Online
- ○ TV
- ○ Radio
- ○ Print news

Which of the newspapers below have you read at least once during the last 12 months? Please select all that apply.

- ☐ Breitbart
- ☐ BuzzFeed News
- ☐ Boston Herald
- ☐ Chicago Sun-Times
- ☐ Daily Mail
- ☐ Drudge Report
- ☐ InfoWars
- ☐ Los Angeles Times
- ☐ New Republic
- ☐ Newsmax
- ☐ New York Daily News
- ☐ New York Post
- ☐ Palmer Report
- ☐ The Denver Post
- ☐ The Huffington Post
- ☐ The Mercury News
- ☐ The New York Times
- ☐ The Wall Street Journal
- ☐ The Washington Post
- ☐ The Washington Times
- ☐ USA Today
- ☐ I have not read any of the newspapers above during the past 12 months

How often do you watch or read politics and economic news?

- ○ Frequently
- ○ Somewhat frequently
- ○ Occasionally
- ○ Rarely
- ○ Never

How interested are you in political and economic news?

- ○ Very interested
- ○ Interested
- ○ Somewhat interested
- ○ Not interested
- ○ Not at all interested

How much do you trust news in general?

- ○ A great deal
- ○ A lot
- ○ A moderate amount
- ○ A little
- ○ None at all

→

What is your gender?

- ○ Male
- ○ Female

What is your age?

Which category best describes your highest level of education?

- ○ Did not graduate high school
- ○ High school degree/GED
- ○ Some college
- ○ 2-year college degree
- ○ 4-year college degree
- ○ Master's degree
- ○ Doctoral degree
- ○ Professional degree (JD, MD, MBA)

Which of the following best describes your race or ethnicity?

- ○ African American/Black
- ○ Asian/Asian American
- ○ Caucasian/White
- ○ Native American, Inuit or Aleut
- ○ Native Hawaiian/Pacific Islander
- ○ Other

Are you of Hispanic, Latino, or Spanish origin?

- Yes
- No

What was your household's gross income in 2023?

- Less than $15,000
- $15,000 to $24,999
- $25,000 to $49,999
- $50,000 to $74,999
- $75,000 to $99,999
- $100,000 to $149,999
- $150,000 to $199,999
- $200,000 or more

In politics, as of today, do you consider yourself a Republican, a Democrat, or an Independent?

- Republican
- Democrat
- Independent

What is your region of residence?

- Northeast (CT, ME, MA, NH, RI, VT, NJ, NY, PA)
- Midwest (IL, IN, MI, OH, WI, IA, KS, MK, MO, NE, ND, SD)
- South (DE, DC, FL, GA, MD, NC, SC, VA, WV, AL, KY, MS, TN, AR, LA, OK, TX)
- West (AZ, CO, ID, NM, MT, UT, NV, WY, AK, CA, HI, OR, WA)

What is your current employment status?

- ◯ Full-time employee
- ◯ Part-time employee
- ◯ Self-employed or small business owner
- ◯ Unemployed and looking for work
- ◯ Student
- ◯ Not in labor force (for example: retired or full-time parent)

Are you liberal or conservative?

- ◯ Very liberal
- ◯ Liberal
- ◯ Neither liberal nor conservative
- ◯ Conservative
- ◯ Very conservative

→

## D.2 Demand Experiment

This study is conducted by researchers from the Frankfurt School of Finance & Management, NHH Norwegian School of Economics, University of Cologne, and MIT. You must be at least 18 years of age to participate in this study. If you do not fulfill these requirements, please do not continue any further.

The goal of this study is to assess an AI-powered news platform, focusing on user experience and content relevance.

Your participation will help understanding people's user experience and improve the platform. We do not anticipate any risks or discomforts from participating in this study.

Your responses to this survey will be anonymous; we will not collect your name or any other personal identifiers. We will only collect your Prolific ID to track your participation in this study, administer potential bonus payments, and invite you to potential future follow-up studies.

You are not allowed to participate in this study more than once. If you experience a technical error or problem, do not try to restart or retake the study. Rather, send us an email with a description of your problem and we will get back to you.

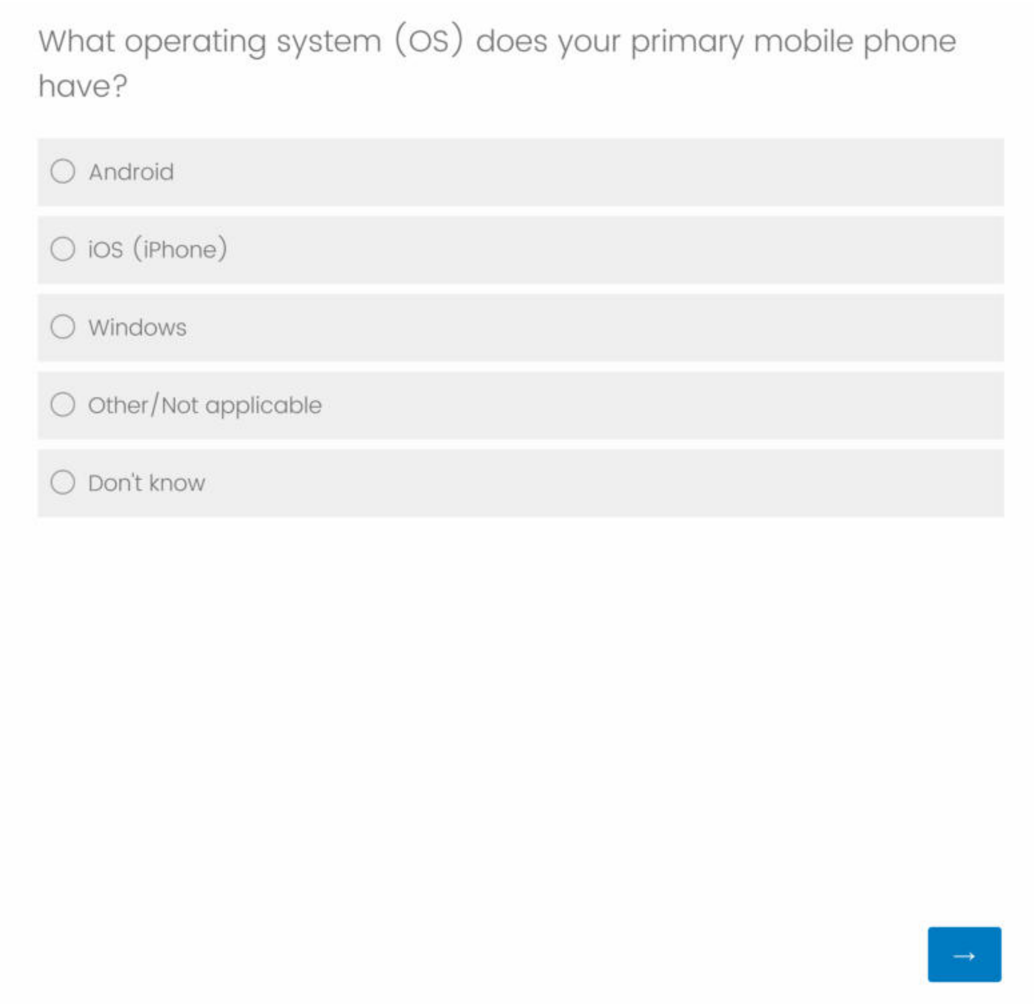
If you have any questions regarding this study, please email roth@wiso.uni-koeln.de

If you feel you have been treated unfairly, or you have questions regarding your rights as a research subject, you may contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, M.I.T., Room E25-143b, 77 Massachusetts Ave, Cambridge, MA 02139, phone 1-617-253-6787.

Do you consent to participate in this study?

- ○ I consent
- ○ I do not consent and wish to exit the study

→


What operating system (OS) does your primary mobile phone have?

- Android
- iOS (iPhone)
- Windows
- Other/Not applicable
- Don't know

→

This question is about the following problem. Sometimes there are participants who do not carefully read the questions and just quickly click through the survey. This means that there are a lot of random answers which compromise the results of research studies. To show that you read our questions carefully, please choose both "Very strongly interested" and "Not at all interested" as your answer to the question below. How interested are you in sports?

- [ ] Very strongly interested
- [ ] Somewhat interested
- [ ] A little interested
- [ ] Almost not interested
- [ ] Not at all interested

→

What is your gender?

- ○ Male
- ○ Female
- ○ Other

What is your age?

Which category best describes your highest level of education?

- ○ Did not graduate high school
- ○ High school degree/GED
- ○ Some college
- ○ 2-year college degree
- ○ 4-year college degree
- ○ Master's degree
- ○ Doctoral degree
- ○ Professional degree (JD, MD, MBA)

Which of the following best describes your race or ethnicity?

- ○ African American/Black
- ○ Asian/Asian American
- ○ Caucasian/White
- ○ Native American, Inuit or Aleut
- ○ Native Hawaiian/Pacific Islander
- ○ Other

Are you of Hispanic, Latino, or Spanish origin?

- ○ Yes
- ○ No

What was your household's gross income (before taxes) in 2023?

- ○ Less than $15,000
- ○ $15,000 to $24,999
- ○ $25,000 to $49,999
- ○ $50,000 to $74,999
- ○ $75,000 to $99,999
- ○ $100,000 to $149,999
- ○ $150,000 to $199,999
- ○ $200,000 or more

In politics, as of today, do you consider yourself a Republican, a Democrat, or an Independent?

- ○ Republican
- ○ Democrat
- ○ Independent

What is your region of residence?

- ○ Northeast (CT, ME, MA, NH, RI, VT, NJ, NY, PA)
- ○ Midwest (IL, IN, MI, OH, WI, IA, KS, MK, MO, NE, ND, SD)
- ○ South (DE, DC, FL, GA, MD, NC, SC, VA, WV, AL, KY, MS, TN, AR, LA, OK, TX)
- ○ West (AZ, CO, ID, NM, MT, UT, NV, WY, AK, CA, HI, OR, WA)

What is your current employment status?

- ○ Full-time employee
- ○ Part-time employee
- ○ Self-employed or small business owner
- ○ Unemployed and looking for work
- ○ Student
- ○ Not in labor force (for example: retired or full-time parent)

Are you liberal or conservative?

- ○ Very liberal
- ○ Liberal
- ○ Neither liberal nor conservative
- ○ Conservative
- ○ Very conservative

Who did you vote for in the 2020 presidential election between Donald Trump and Joe Biden, or did you not vote?

- ○ Donald Trump
- ○ Joe Biden
- ○ Someone else
- ○ Did not vote

Who did you vote for in the 2024 presidential election between Donald Trump and Kamala Harris, or do you not vote?

- ○ Donald Trump
- ○ Kamala Harris
- ○ Someone else
- ○ Did not vote

→

Do you currently have any paid subscriptions to news platforms (printed newspaper, news website or app, newsletter)?

- ◯ Yes
- ◯ No

How do you primarily consume news?

- ◯ Online
- ◯ TV
- ◯ Radio
- ◯ Print news

For reading news, what device do you mainly use?

- ◯ Phone
- ◯ Tablet
- ◯ Computer
- ◯ Print
- ◯ I don't read any news

How many news apps do you have on your phone?

→

How satisfied are you with your overall news consumption over the past few days?

- ◯ Very satisfied
- ◯ Somewhat satisfied
- ◯ Neutral
- ◯ Somewhat dissatisfied
- ◯ Very dissatisfied

How interested are you in political and economic news?

- ◯ Very interested
- ◯ Interested
- ◯ Somewhat interested
- ◯ Not interested
- ◯ Not at all interested

How often do you watch or read politics and economic news?

- ◯ Frequently
- ◯ Somewhat frequently
- ◯ Occasionally
- ◯ Rarely
- ◯ Never

How often do you feel fatigued or tired of hearing about political and economic news?

- ◯ Always
- ◯ Often
- ◯ Sometimes
- ◯ Rarely
- ◯ Never

How satisfied are you with how the news media currently covers political and economic news?

- ◯ Very satisfied
- ◯ Somewhat satisfied
- ◯ Neither satisfied nor dissatisfied
- ◯ Somewhat dissatisfied
- ◯ Very dissatisfied

→

In this survey, we are going to tell you about our app and will ask whether you are interested in downloading it to your phone. **Please note that your payment for the survey will not depend on whether you download the app.** Everyone who finishes the survey will be paid irrespective of whether they download the app.

→

**Get early access to our news app!**

We're excited to offer you access to our AI-powered news app.

TAILORMADETIMES

**Our sources**
Each article is rewritten using advanced AI technology to deliver essential information quickly and clearly. The content is based on trending stories from the Associated Press (AP), ensuring professional journalism and high accuracy.

**Accurate reporting**
While our AI tailors the presentation to make it more concise, the underlying facts mentioned in the article remain unchanged from the AP sources, guaranteeing the same level of factual accuracy. We also have a human review of articles, just to be sure.

One the next page, you will see what the app looks like.

→

**Get early access to our news app!**

We're excited to offer you access to our AI-powered news app.

TAILORMADETIMES

**Tailored news, vetted facts!**

The key feature of our news app is that you can tailor the news stories according to your preferences. Click the **Tailor** button and move the sliders to choose how you want the news to be reported:

- **Complexity** *(regular, less complex)*: Makes the article more accessible so readers can understand the content easily and quickly. Replaces difficult terms with simpler ones and adds brief explanations for unfamiliar concepts.
- **Sentiment** *(more positive, regular, more negative)*: Changes the language of the article to be either more positive or negative. Highlights positive aspects and avoids negative terms to create a more positive sentiment, or emphasizes potential downsides for a more negative sentiment.
- **Entertainment** *(regular, entertaining)*: Makes the article more entertaining by adding humor, puns, and clever word choices. Also rewrites some articles using a playful narrative style.
- **Political learning** *(liberal, regular, and conservative)*: Changes the political perspective of the article by using language and terms that are more commonly used by either conservatives or liberals. Also changes the focus of the article to highlight themes and issues that matter most to each political group.

You can change slider settings at any point in time, and the app will update the writing of all articles in real-time. **Irrespective of how you tailor the news, the underlying facts remain constant.** Tailoring does also not affect the selection of articles - it only changes how they are presented.

**Our sources**

Each article is rewritten using advanced AI technology to deliver essential information quickly and clearly. The content is based on trending stories from the Associated Press (AP), ensuring professional journalism and high accuracy.

**Accurate reporting**

While our AI tailors the presentation to make it more concise and align it with your preferences, the underlying facts mentioned in the article remain unchanged from the AP sources, guaranteeing the same level of factual accuracy. We also have a human review of articles, just to be sure.

One the next page, you will see what the app looks like.

**Get early access to our news app!**

We're excited to offer you access to our AI-powered news app.

TAILORMADETIMES

**Tailored news, vetted facts!**

The key feature of our news app is that you can tailor the news stories according to your preferences. Click the **Tailor** button and move the sliders to choose how you want the news to be reported:

- **Complexity** *(regular, less complex)*: Makes the article more accessible so readers can understand the content easily and quickly. Replaces difficult terms with simpler ones and adds brief explanations for unfamiliar concepts.
- **Sentiment** *(more positive, regular, more negative)*: Changes the language of the article to be either more positive or negative. Highlights positive aspects and avoids negative terms to create a more positive sentiment, or emphasizes potential downsides for a more negative sentiment.
- **Entertainment** *(regular, entertaining)*: Makes the article more entertaining by adding humor, puns, and clever word choices. Also rewrites some articles using a playful narrative style.
- **Facts only** *(regular, facts only)*: Makes the article less opinionated by focusing on facts only. Also avoids subjective assessments and judgements, and uses more neutral language.

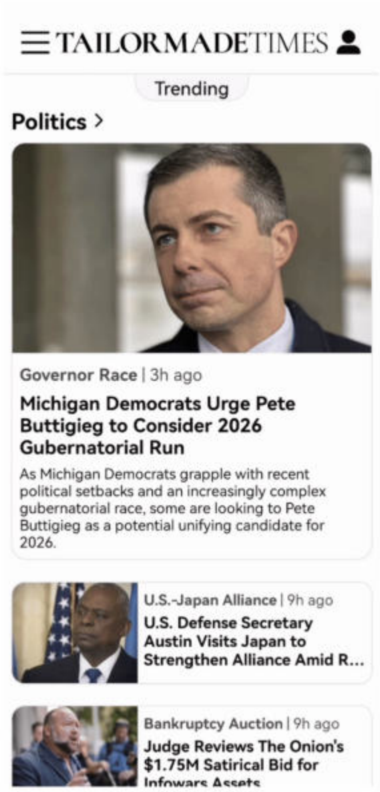
You can change slider settings at any point in time, and the app will update the writing of all articles in real-time. **Irrespective of how you tailor the news, the underlying facts remain constant.** Tailoring does also not affect the selection of articles - it only changes how they are presented.

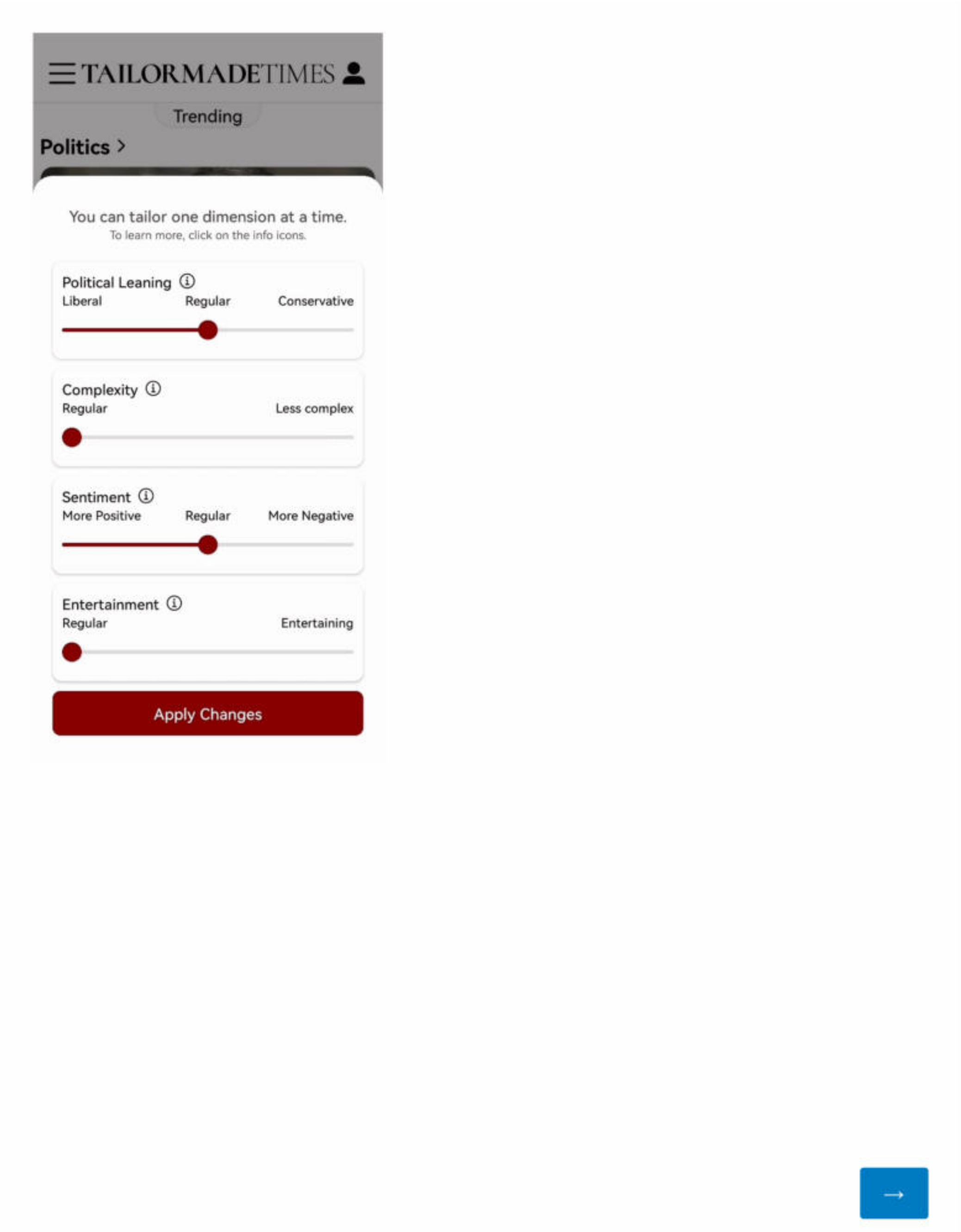
**Our sources**
Each article is rewritten using advanced AI technology to deliver essential information quickly and clearly. The content is based on trending stories from the Associated Press (AP), ensuring professional journalism and high accuracy.

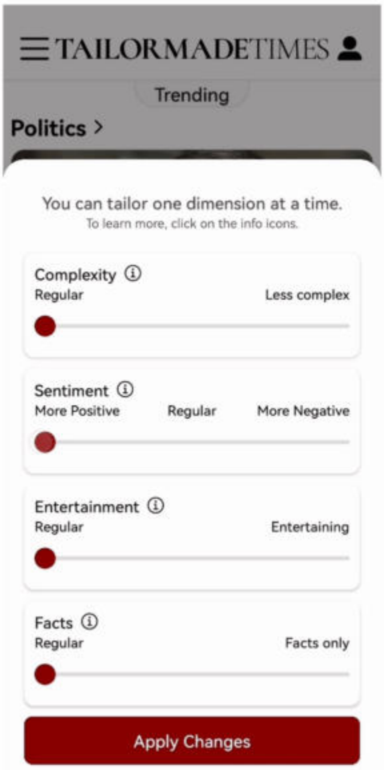
**Accurate reporting**
While our AI tailors the presentation to make it more concise and align it with your preferences, the underlying facts mentioned in the article remain unchanged from the AP sources, guaranteeing the same level of factual accuracy. We also have a human review of articles, just to be sure.

One the next page, you will see what the app looks like.

→


TAILORMADETIMES
Trending
Politics >
Governor Race | 3h ago
Michigan Democrats Urge Pete Buttigieg to Consider 2026 Gubernatorial Run
As Michigan Democrats grapple with recent political setbacks and an increasingly complex gubernatorial race, some are looking to Pete Buttigieg as a potential unifying candidate for 2026.
U.S.-Japan Alliance | 9h ago
U.S. Defense Secretary Austin Visits Japan to Strengthen Alliance Amid R...
Bankruptcy Auction | 9h ago
Judge Reviews The Onion's $1.75M Satirical Bid for


TAILORMADETIMES
Trending
Politics >
You can tailor one dimension at a time.
To learn more, click on the info icons.
Political Leaning
Liberal
Regular
Conservative
Complexity
Regular
Less complex
Sentiment
More Positive
Regular
More Negative
Entertainment
Regular
Entertaining
Apply Changes




You will now receive the opportunity to download the news app to your phone. If you say yes, we will provide you with instructions on how to download the app on the next page.

- ◯ Yes, I want to download the app
- ◯ No, I do not want to download the app

→

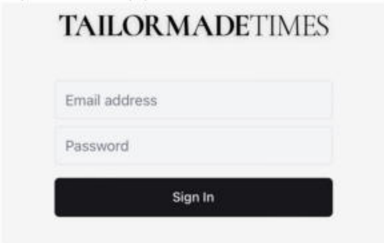
## How to download the news app before its public release

Below are instructions how you can get **exclusive access** to the news app **before its public release**.

**Step 1: Download our app *TailorMadeTimes***
Download our app using this link: https://play.google.com/store/apps/details?id=com.tailormadetimes.app or scanning this QR:

SCAN ME

**Step 2: Open the app *TailorMadeTimes***
Open the app. You will see this:




**Step 3: Log in**
Log in using your login details below:

Username: [redacted] Copy Username
Password: [redacted] Copy Password

We will also send you your username and password as a message on Prolific in case you need to look them up again later.

→

**How to download the news app before its public release**

Below are instructions how you can get **exclusive access** to the news app **before its public release**.

**Step 1: Download the app *TestFlight***
TestFlight is an app developed by Apple that allows developers to distribute beta versions of their apps to a select group of testers before the official release. All apps in TestFlight are carefully reviewed and approved by Apple.

Download the app from the the App Store using this link: https://apps.apple.com/us/app/testflight/id899247664 or scanning this QR:

SCAN ME

Then continue with Step 2 to get our app (***you cannot enter a code in TestFlight directly!***)

**Step 2: Download our app *TailorMadeTimes***
Download our app using this link: https://testflight.apple.com/join/w6lyC7Zg or scanning this QR:

SCAN ME

**Step 3: Open the app *TailorMadeTimes***
Open the app. You will see this:


**Step 4: Log in**
Log in using your login details below:

Username: Copy Username
Password: Copy Password

We will also send you your username and password as a message on Prolific in case you need to look them up again later.

→

If you plan to download the app, it is important that you follow the instructions on the previous page. You will not be able to download the app at a later stage.

Do you want to see the instructions again on how to download the app?

- ◯ Yes, I want to see the instructions again
- ◯ No, I have already downloaded the app or no longer want to download the app

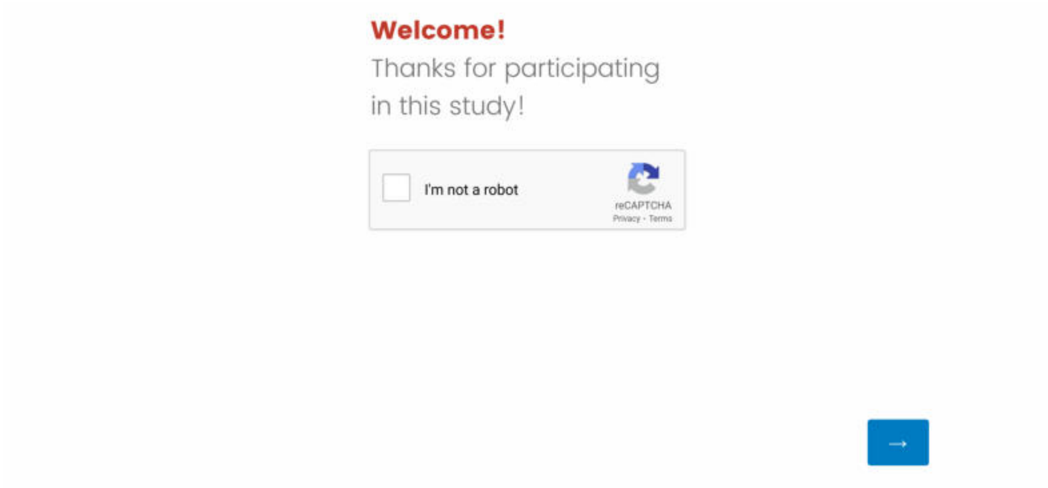
→

Did you download the app?

- ◯ Yes
- ◯ No

→

## D.3 Usage Incentive Experiment



This study is conducted by researchers from the Frankfurt School of Finance & Management, NHH Norwegian School of Economics, University of Cologne, and MIT. You must be at least 18 years of age to participate in this study. If you do not fulfill these requirements, please do not continue any further.

The goal of this study is to assess an AI-tailored news platform, focusing on user experience and content relevance.

Your participation will help understanding people's user experience and improve the platform. We do not anticipate any risks or discomforts from participating in this study.

Your responses to this survey will be anonymous; we will not collect your name or any other personal identifiers. We will only collect your Prolific ID to track your participation in this study, administer potential bonus payments, and invite you to potential future follow-up studies.

You are not allowed to participate in this study more than once. If you experience a technical error or problem, do not try to restart or retake the study. Rather, send us an email with a description of your problem and we will get back to you.

If you have any questions regarding this study, please email roth@wiso.uni-koeln.de

If you feel you have been treated unfairly, or you have questions regarding your rights as a research subject, you may contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, M.I.T., Room E25-143b, 77 Massachusetts Ave, Cambridge, MA 02139, phone 1-617-253-6787.

Do you consent to participate in this study?

○ I consent

○ I do not consent and wish to exit the study

→

This question is about the following problem. Sometimes there are participants who do not carefully read the questions and just quickly click through the survey. This means that there are a lot of random answers which compromise the results of research studies. To show that you read our questions carefully, please choose both "Very strongly interested" and "Not at all interested" as your answer to the question below. How interested are you in sports?

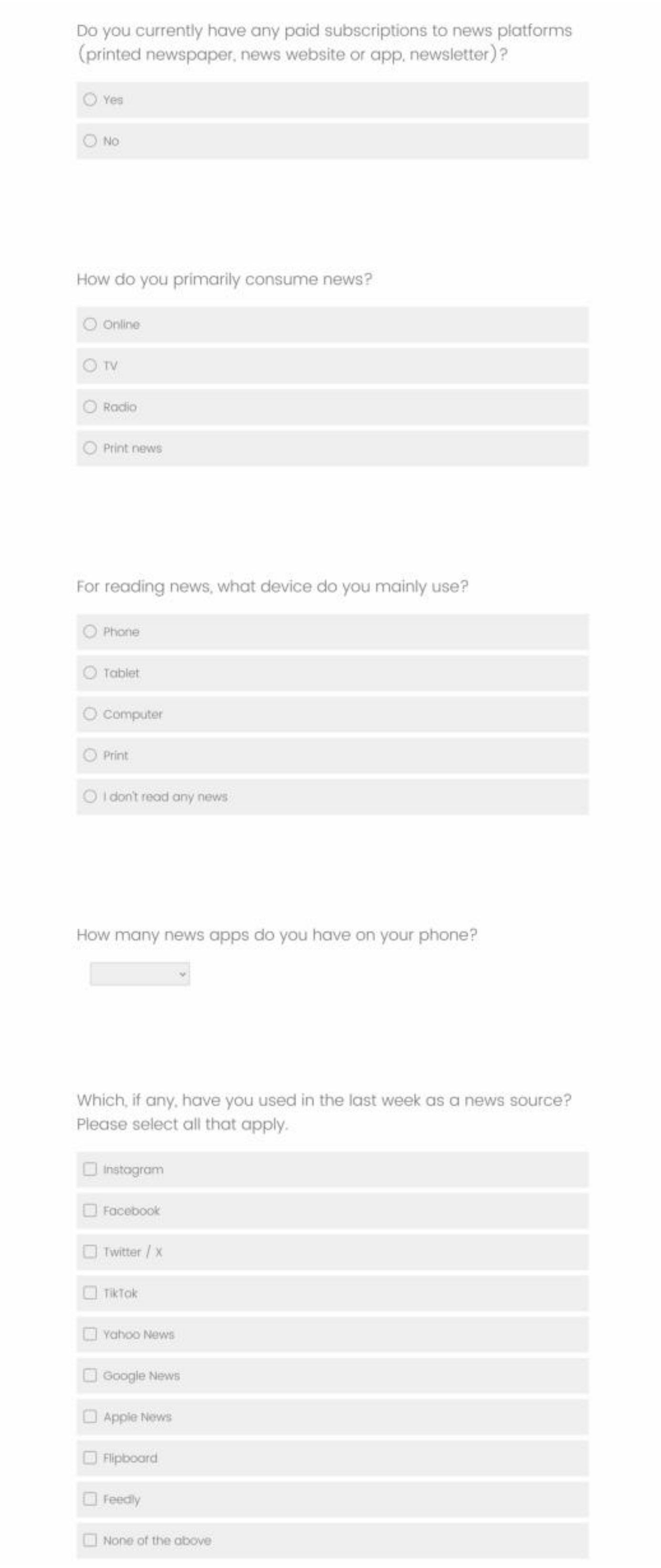
- [ ] Very strongly interested
- [ ] Somewhat interested
- [ ] A little interested
- [ ] Almost not interested
- [ ] Not at all interested

→

What operating system (OS) does your primary mobile phone have?

- ○ Android
- ○ iOS (iPhone)
- ○ Windows
- ○ Other/Not applicable
- ○ Don't know

→



Do you currently have any paid subscriptions to news platforms (printed newspaper, news website or app, newsletter)?

- Yes
- No

How do you primarily consume news?

- Online
- TV
- Radio
- Print news

For reading news, what device do you mainly use?

- Phone
- Tablet
- Computer
- Print
- I don't read any news

How many news apps do you have on your phone?

Which, if any, have you used in the last week as a news source? Please select all that apply.

- Instagram
- Facebook
- Twitter / X
- TikTok
- Yahoo News
- Google News
- Apple News
- Flipboard
- Feedly
- None of the above

Which of the news outlets below have you consumed at least once during the last 12 months? Please select all that apply.

- ☐ Breitbart
- ☐ BuzzFeed News
- ☐ Boston Herald
- ☐ Chicago Sun-Times
- ☐ Daily Mail
- ☐ Drudge Report
- ☐ InfoWars
- ☐ Los Angeles Times
- ☐ New Republic
- ☐ Newsmax
- ☐ New York Daily News
- ☐ New York Post
- ☐ Palmer Report
- ☐ The Denver Post
- ☐ The Huffington Post
- ☐ The Mercury News
- ☐ The New York Times
- ☐ The Wall Street Journal
- ☐ The Washington Post
- ☐ The Washington Times
- ☐ USA Today
- ☐ I have not consumed any of the news outlets above during the past 12 months

→

How satisfied are you with your overall news consumption over the past few days?

- ○ Very satisfied
- ○ Somewhat satisfied
- ○ Neutral
- ○ Somewhat dissatisfied
- ○ Very dissatisfied

How interested are you in political and economic news?

- ○ Very interested
- ○ Interested
- ○ Somewhat interested
- ○ Not interested
- ○ Not at all interested

How often do you watch or read politics and economic news?

- ○ Frequently
- ○ Somewhat frequently
- ○ Occasionally
- ○ Rarely
- ○ Never

How often do you feel fatigued or tired of hearing about political and economic news?

- ○ Always
- ○ Often
- ○ Sometimes
- ○ Rarely
- ○ Never

How satisfied are you with how the news media currently covers political and economic news?

- ◯ Very satisfied
- ◯ Somewhat satisfied
- ◯ Neither satisfied nor dissatisfied
- ◯ Somewhat dissatisfied
- ◯ Very dissatisfied

How much do you agree or disagree with the following statements?

| | Strongly disagree | Somewhat disagree | Neutral | Somewhat agree | Strongly agree |
|---|---|---|---|---|---|
| News articles are too **complex** | ◯ | ◯ | ◯ | ◯ | ◯ |
| News articles are too **negative** | ◯ | ◯ | ◯ | ◯ | ◯ |
| News articles are too **positive** | ◯ | ◯ | ◯ | ◯ | ◯ |
| News articles are **not entertaining enough** | ◯ | ◯ | ◯ | ◯ | ◯ |
| News articles are too **liberal** | ◯ | ◯ | ◯ | ◯ | ◯ |
| News articles are too **conservative** | ◯ | ◯ | ◯ | ◯ | ◯ |
| News articles are too **opinionated** | ◯ | ◯ | ◯ | ◯ | ◯ |

→

How much do you trust news in general?

- A great deal
- A lot
- A moderate amount
- A little
- None at all

How much do you trust news about politics and the economy?

- A great deal
- A lot
- A moderate amount
- A little
- None at all

→

How much do you trust AI technology, such as language models, in general?

- ◯ A great deal
- ◯ A lot
- ◯ A moderate amount
- ◯ A little
- ◯ None at all

How much do you trust AI to accurately and impartially tailor news about politics and the economy?

- ◯ A great deal
- ◯ A lot
- ◯ A moderate amount
- ◯ A little
- ◯ None at all

→

How often do you use ChatGPT or other large language models?

- Every day
- Multiple times per week
- A few times a month
- Once a month
- Never

→

To what extent do you agree with the following statement: *AI technology has the potential to make the news more relevant for me.*

- Strongly agree
- Somewhat agree
- Neither agree not disagree
- Somewhat disagree
- Strongly disagree

→

We would like you to rate your feelings toward some different groups and individuals with whom you may agree or disagree politically. Ratings are made on what we call a 'feeling thermometer.' The scale ranges from 0 to 100. When you rate an individual or group, a score of 0 degrees means you feel as cold and unfavorable as possible towards them. A score of 100 degrees means you feel as warm and favorable as possible. You would rate them at 50 degrees if you do not feel particularly warm or cold toward them.

**Please rate how you feel about the following groups or individuals.**

| Cold and unfavorable | | | | | | | | | | Warm and favorable |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |

**Democrats**

**Republicans**

→

Imagine that you had to split 100 dollars between two other people, Person A and Person B. Person A is a randomly chosen person among those who voted for the Democrat party in the last presidential election, while Person B is a randomly chosen person among those who voted for the Republican party in the last presidential election.

How much money would you like to give to Person A and how much money would you like to give to Person B?

| | |
|---|---|
| **Person A** (a Democrat) | 0 |
| **Person B** (a Republican) | 0 |
| Total | 0 |

→

Tell me whether you have a very favorable, somewhat favorable, somewhat unfavorable or very unfavorable opinion of Donald Trump.

- ○ Very favorable
- ○ Somewhat favorable
- ○ Somewhat unfavorable
- ○ Very unfavorable

Tell me whether you have a very favorable, somewhat favorable, somewhat unfavorable or very unfavorable opinion of Joe Biden.

- ○ Very favorable
- ○ Somewhat favorable
- ○ Somewhat unfavorable
- ○ Very unfavorable

Tell me whether you have a very favorable, somewhat favorable, somewhat unfavorable or very unfavorable opinion of Kamala Harris.

- ○ Very favorable
- ○ Somewhat favorable
- ○ Somewhat unfavorable
- ○ Very unfavorable

→

Overall, do you think that free trade agreements between the US and other countries have been a good thing or a bad thing for the United States?

- ○ Good thing
- ○ Bad thing

Overall, would you say that blacks or whites are treated more fairly in dealing with the police?

- ○ Blacks
- ○ Whites

Do you think that employers firing men who have been accused of sexual harassment or assault before finding out all the facts is a major or a minor problem?

- ○ Major problem
- ○ Minor problem

On the whole, do you think immigration is a good thing or a bad thing for this country today?

- ○ Good thing
- ○ Bad thing

In general, do you feel that the laws covering the sale of firearms should be made less strict, more strict, or kept as they are now?

- ○ Less strict
- ○ Kept as they are
- ○ More strict

→

Over the next presidential turn, how much higher or lower do you expect the average **inflation rate** to be under a Trump presidency compared to a Harris presidency?

- ○ More than 10 percentage points higher under Trump
- ○ Between 5 and 10 percentage points higher under Trump
- ○ Between 2 and 5 percentage points higher under Trump
- ○ Between 0 and 2 percentage points higher under Trump
- ○ Between 0 and 2 percentage points lower under Trump
- ○ Between 2 and 5 percentage points lower under Trump
- ○ Between 5 and 10 percentage points lower under Trump
- ○ More than 10 percentage points lower under Trump

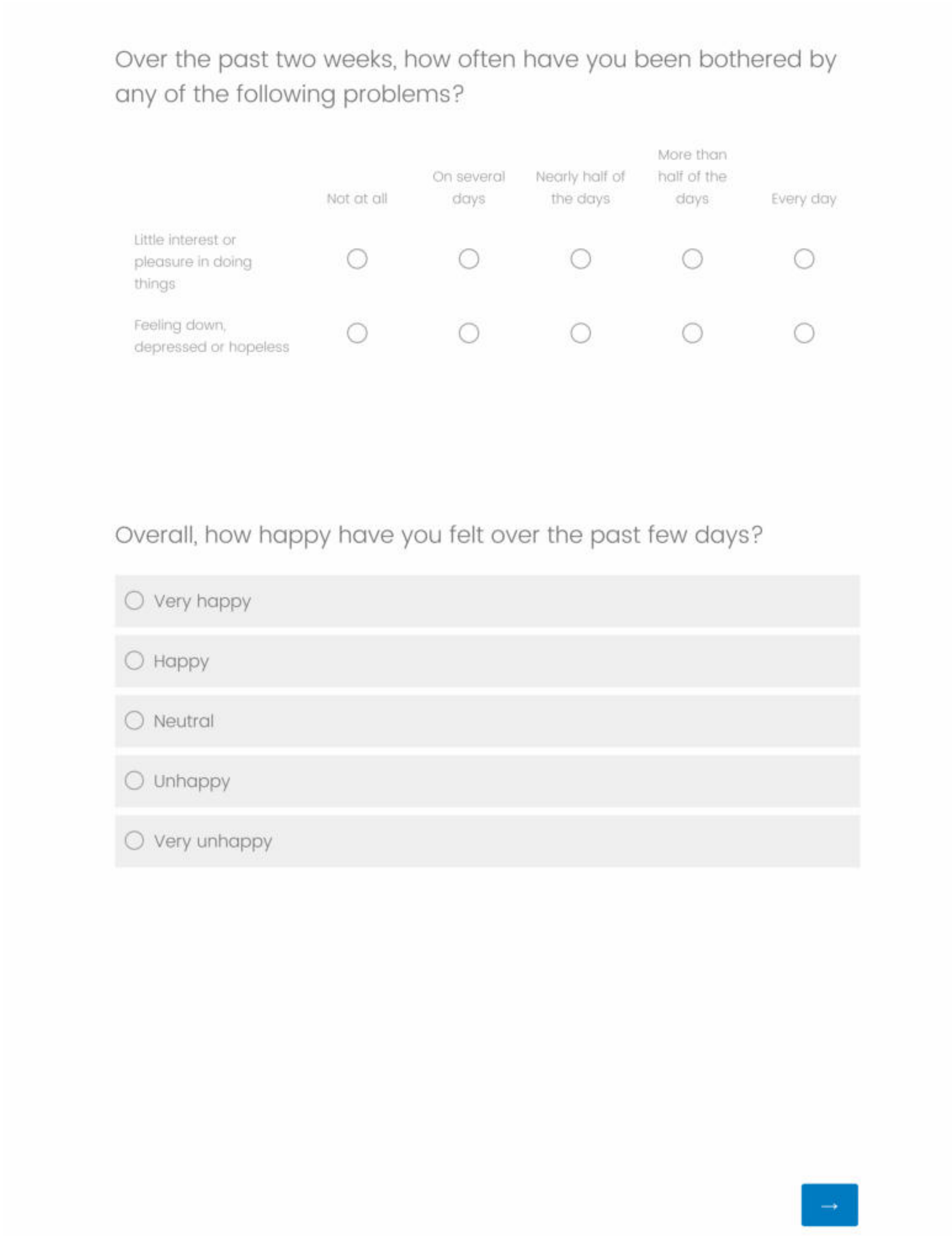
Over the next presidential turn, how much higher or lower do you expect the average **crime rate** to be under a Trump presidency compared to a Harris presidency?

- ○ More than 10 percentage points higher under Trump
- ○ Between 5 and 10 percentage points higher under Trump
- ○ Between 2 and 5 percentage points higher under Trump
- ○ Between 0 and 2 percentage points higher under Trump
- ○ Between 0 and 2 percentage points lower under Trump
- ○ Between 2 and 5 percentage points lower under Trump
- ○ Between 5 and 10 percentage points lower under Trump
- ○ More than 10 percentage points lower under Trump

Over the next presidential turn, how much higher or lower do you expect the average **unemployment rate** to be under a Trump presidency compared to a Harris presidency?

- ◯ More than 10 percentage points higher under Trump
- ◯ Between 5 and 10 percentage points higher under Trump
- ◯ Between 2 and 5 percentage points higher under Trump
- ◯ Between 0 and 2 percentage points higher under Trump
- ◯ Between 0 and 2 percentage points lower under Trump
- ◯ Between 2 and 5 percentage points lower under Trump
- ◯ Between 5 and 10 percentage points lower under Trump
- ◯ More than 10 percentage points lower under Trump

→



Over the past two weeks, how often have you been bothered by any of the following problems?

| | Not at all | On several days | Nearly half of the days | More than half of the days | Every day |
|---|---|---|---|---|---|
| Little interest or pleasure in doing things | ○ | ○ | ○ | ○ | ○ |
| Feeling down, depressed or hopeless | ○ | ○ | ○ | ○ | ○ |

Overall, how happy have you felt over the past few days?

- ○ Very happy
- ○ Happy
- ○ Neutral
- ○ Unhappy
- ○ Very unhappy

→

**Early access to an AI-tailored news app**

Be among the first to experience our new AI-tailored news app for iPhone and Android. We are offering exclusive **free early access** to a select group of users before the official launch. You have the chance to explore the site and discover its features ahead of the public.

On the next page, we will provide more information about the news app.

→

## Transforming news with AI technology

Before deciding on whether you want **free early access**, we want to highlight some unique aspects of our **AI-tailored news app:**

### Stay informed with the most relevant news

Experience the latest and most relevant news stories from the US. Our AI-tailored news platform provides frequent updates, ensuring you stay informed with the most current and trending stories without being overwhelmed by lengthy articles.

### Concise and accurate articles

Each article is rewritten using advanced AI technology to deliver essential information quickly and clearly. The content is based on trending stories from the **Associated Press (AP)**, ensuring professional journalism and high accuracy.

### Trustworthy reporting

While our AI improves the presentation, the core facts remain unchanged from the AP sources, guaranteeing the same level of trust and credibility you expect from seasoned journalists.

→

**Are you interested in early access to our news app?**

Please tell us whether you are interested in getting early access to our news app.

- ◯ Yes, I want early access
- ◯ No, I decline early access

→

We will invite you for a follow-up survey in the next few days where we'll provide instructions on how to get early access.

→

What is your gender?

- ○ Male
- ○ Female
- ○ Other

What is your age?

Which category best describes your highest level of education?

- ○ Did not graduate high school
- ○ High school degree/GED
- ○ Some college
- ○ 2-year college degree
- ○ 4-year college degree
- ○ Master's degree
- ○ Doctoral degree
- ○ Professional degree (JD, MD, MBA)

Which of the following best describes your race or ethnicity?

- ○ African American/Black
- ○ Asian/Asian American
- ○ Caucasian/White
- ○ Native American, Inuit or Aleut
- ○ Native Hawaiian/Pacific Islander
- ○ Other

Are you of Hispanic, Latino, or Spanish origin?

- ○ Yes
- ○ No

What was your household's gross income (before taxes) in 2023?

- Less than $15,000
- $15,000 to $24,999
- $25,000 to $49,999
- $50,000 to $74,999
- $75,000 to $99,999
- $100,000 to $149,999
- $150,000 to $199,999
- $200,000 or more

In politics, as of today, do you consider yourself a Republican, a Democrat, or an Independent?

- Republican
- Democrat
- Independent

What is your region of residence?

- Northeast (CT, ME, MA, NH, RI, VT, NJ, NY, PA)
- Midwest (IL, IN, MI, OH, WI, IA, KS, MK, MO, NE, ND, SD)
- South (DE, DC, FL, GA, MD, NC, SC, VA, WV, AL, KY, MS, TN, AR, LA, OK, TX)
- West (AZ, CO, ID, NM, MT, UT, NV, WY, AK, CA, HI, OR, WA)

What is your current employment status?

- Full-time employee
- Part-time employee
- Self-employed or small business owner
- Unemployed and looking for work
- Student
- Not in labor force (for example: retired or full-time parent)

Are you liberal or conservative?

- ◯ Very liberal
- ◯ Liberal
- ◯ Neither liberal nor conservative
- ◯ Conservative
- ◯ Very conservative

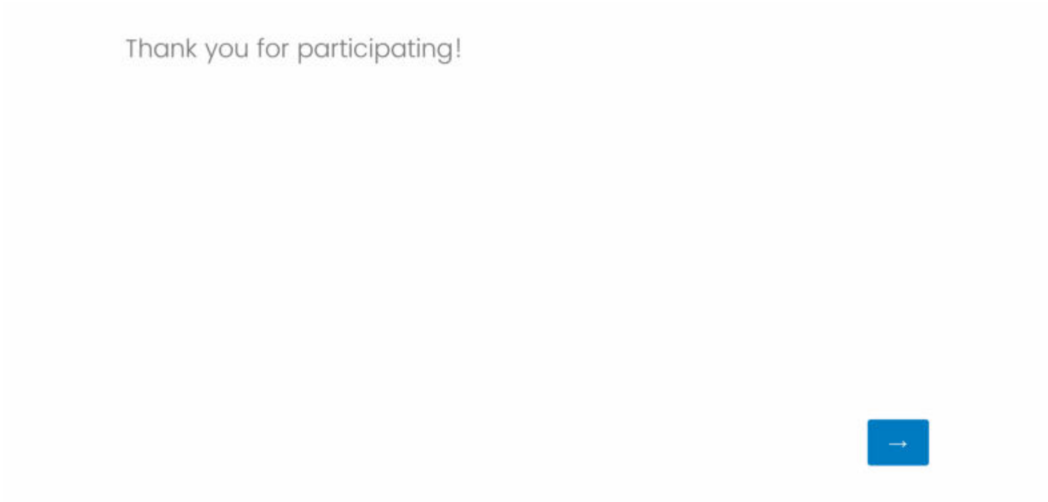
Who did you vote for in the 2020 presidential election between Donald Trump and Joe Biden, or did you not vote?

- ◯ Donald Trump
- ◯ Joe Biden
- ◯ Someone else
- ◯ Did not vote

Who did you vote for in the 2024 presidential election between Donald Trump and Kamala Harris, or did you not vote?

- ◯ Donald Trump
- ◯ Kamala Harris
- ◯ Someone else
- ◯ Did not vote

→


Thank you for participating!
→

### D.3.1 Download Survey

This study is conducted by researchers from the Frankfurt School of Finance & Management, NHH Norwegian School of Economics, University of Cologne, and MIT. You must be at least 18 years of age to participate in this study. If you do not fulfill these requirements, please do not continue any further.

The goal of this study is to assess an AI-tailored news platform, focusing on user experience and content relevance.

Your participation will help understanding people's user experience and improve the platform. We do not anticipate any risks or discomforts from participating in this study.

Your responses to this survey will be anonymous; we will not collect your name or any other personal identifiers. We will only collect your Prolific ID to track your participation in this study, administer potential bonus payments, and invite you to potential future follow-up studies.

You are not allowed to participate in this study more than once. If you experience a technical error or problem, do not try to restart or retake the study. Rather, send us an email with a description of your problem and we will get back to you.

If you have any questions regarding this study, please email roth@wiso.uni-koeln.de

If you feel you have been treated unfairly, or you have questions regarding your rights as a research subject, you may contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, M.I.T., Room E25-143b, 77 Massachusetts Ave, Cambridge, MA 02139, phone 1-617-253-6787.

Do you consent to participate in this study?

- ○ I consent
- ○ I do not consent and wish to exit the study

→

## Early access to an AI-tailored news app

You recently completed our survey and expressed interest in our AI-tailored news app.

Are you still interested in getting early access to the app?

- ◯ Yes, I'm interested
- ◯ No, I'm not interested

→

**Get early access to our AI-tailored news app and earn a $2 bonus!**

We're excited to offer you early access to our AI-tailored news app, ***TailorMadeTimes***, available to you before its public launch.

TAILORMADETIMES

Downloading the app is optional and will not impact your base payment for completing this survey.

If you download the app and enable push notifications, you will receive a $2 bonus.

## How do you get the bonus?

In order to qualify for the $2 bonus you will need to complete the following steps:

- Download our app from the Google Play Store
- Use the login credentials that we provide you with to log into our app
- Start a short onboarding process in our app to receive a verification code
- Enter the verification code in the survey
- Complete the onboarding process on the app and enable push notifications

**Are you willing to download the app and enable push notifications?**

(If you are willing to download the app and enable push notifications, you will get additional information on how to get the app and complete the above steps.)

- ○ Yes, I am willing to download the app and enable push notifications for the $2 bonus
- ○ No, I am not interested in downloading the app

→

**How do you get the bonus?**

In order to qualify for the $2 bonus you will need to complete the following steps:

- Download TestFlight on your iPhone (an offical app from Apple)
- Download our app through TestFlight
- Use the login credentials that we provide you with to log into our app
- Start a short onboarding process in our app to receive a verification code
- Enter the verification code in the survey
- Complete the onboarding process on the app and enable push notifications

**Are you willing to download the app and enable push notifications?**

(If you are willing to download the app and enable push notifications, you will get additional information on how to get the app and complete the above steps.)

- ◯ Yes, I am willing to download the app and enable push notifications for the $2 bonus
- ◯ No, I am not interested in downloading the app

→

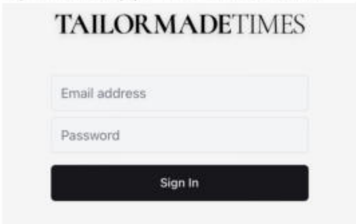
**How to download the news app before its public release**

Below are instructions how you can get **exclusive access** to the news app **before its public release**.

**Step 1: Download our app** ***TailorMadeTimes***
Download our app using this link: https://play.google.com/store/apps/details?id=com.tailormadetimes.app or scanning this QR:

SCAN ME

**Step 2: Open the app** ***TailorMadeTimes***
Open the app. You will see this:

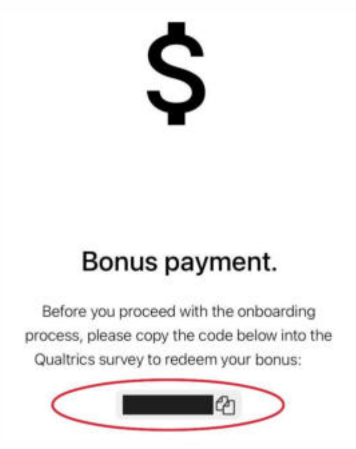



**Step 3: Log in**
Log in using your login details below:

Username: Copy Username
Password: Copy Password

We will also send you your username and password as a message on Prolific in case you need to look them up again later.

**Step 4: Enter your access code**
You will see this app onboarding screen with an access code.



**Enter the access code here:**

*You only qualify for the bonus if you enter the access code.*

After you have copy pasted the access code from the app to the survey you can proceed to the next page in the app which allows you to enable push notifications.

To receive the $2 bonus, you will then be asked to turn on push notifications for the TailorMadeTimes app. Moving the slider to the right allows you to turn on push notifications in your phone's settings and after that you will be able to switch the slider to green.


Did you enable push notifications for TailorMadeTimes?

- ◯ Yes
- ◯ No

→

**How do you download the app**

**Step 1: Download the app *TestFlight***
TestFlight is an app developed by Apple that allows developers to distribute beta versions of their apps to a select group of testers before the official release. All apps in TestFlight are carefully reviewed and approved by Apple.

Download the app from the the App Store using this link: https://apps.apple.com/us/app/testflight/id899247664 or scanning this QR code:

SCAN ME

Then continue with Step 2 to get our app (**you cannot enter a code in TestFlight directly!**)

**Step 2: Download our app *TailorMadeTimes* through *TestFlight***
Now you can download our app though TestFlight using this link: https://testflight.apple.com/join/w6lyC7Zg or scanning this QR:

SCAN ME

**Step 3: Log in to the app**
Open the app. You will see this:

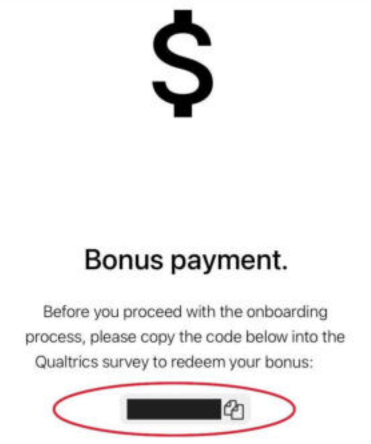

**Step 4: Log in**
Log in using your login details below:

Username: Copy Username
Password: Copy Password

We will also send you your username and password as a message on Prolific in case you need to look them up again later.

**Step 5: Enter your access code**

You will see this app onboarding screen with an access code.




**Enter the access code here:**

*You only qualify for the bonus if you enter the access code.*

After you have copy pasted the access code from the app to the survey you can proceed to the next page in the app which allows you to enable push notifications.

To receive the $2 bonus, you will then be asked to turn on push notifications for the TailorMadeTimes app. Moving the slider to the right allows you to turn on push notifications in your phone's settings and after that you will be able to switch the slider to green.

Did you enable push notifications for TailorMadeTimes?

- ◯ Yes
- ◯ No

→

## Onboarding

Please take a few minutes to finish the onboarding process in the app and read one or two articles to get a sense of how the app works.

We will then ask you a few questions about your user experience.

When you're done with the onboarding and have read one or two articles, click next.

→

## Feedback

We'd like your feedback about the onboarding and your first impression of the app.

Did you finish the onboarding in the app?

- ○ Yes
- ○ No

How many articles did you read?

- ○ 0
- ○ 1
- ○ 2
- ○ 3
- ○ More than 3

Please give us some feedback about different aspects of the app. For each aspect, tell us how much you liked it on a scale form 0 (Do not like it at all) to 10 (Like it a lot).

0 1 2 3 4 5 6 7 8 9 10

Layout

Ease of use

Content

Onboarding

How was your experience with downloading and setting up the app? Please let us know about any difficulties or challenges you encountered.

→

Did you use the **tailor** button to customize articles?

- ◯ Yes
- ◯ No

→

Which customization dimensions did you use? Check all that apply.

- ☐ Political leaning: liberal
- ☐ Political leaning: conservative
- ☐ Less complex
- ☐ Sentiment: more positive
- ☐ Sentiment: more negative
- ☐ Entertaining

→

## Some more information about the app

### Our sources

Each article is rewritten using advanced AI technology to deliver essential information quickly and clearly. The content is based on trending stories from the Associated Press (AP), ensuring professional journalism and high accuracy.

### Accurate reporting

While our AI improves the presentation, the underlying facts mentioned in the article remain unchanged from the AP sources, guaranteeing the same level of factual accuracy. We also have a human review of articles, just to be sure.

→

**Some more information about the app**

**Our sources**
Each article is rewritten using advanced AI technology to deliver essential information quickly and clearly. The content is based on trending stories from the Associated Press (AP), ensuring professional journalism and high accuracy.

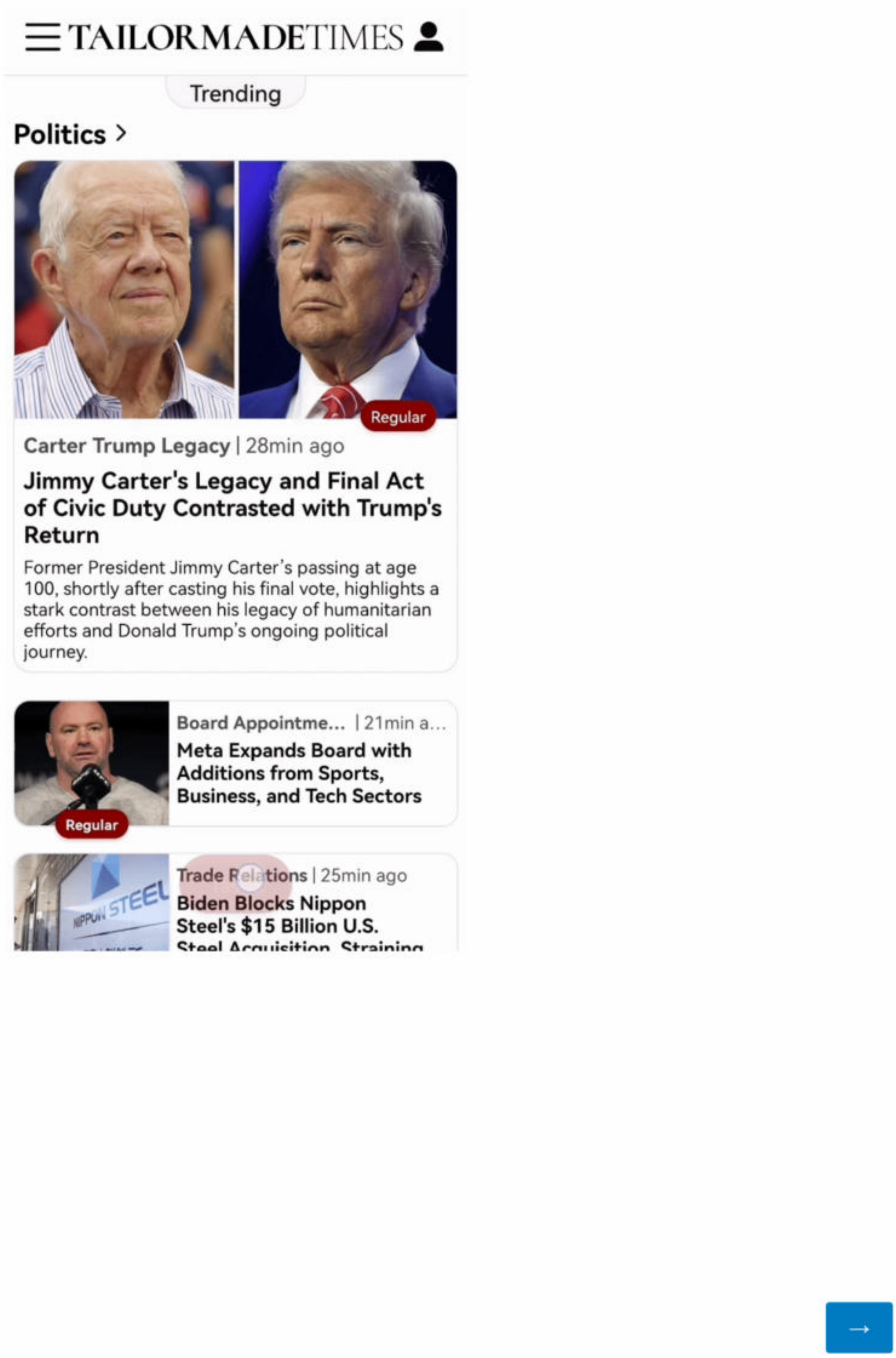
**Accurate reporting**
While our AI improves the presentation, the underlying facts mentioned in the article remain unchanged from the AP sources, guaranteeing the same level of factual accuracy. We also have a human review of articles, just to be sure.

**Customization**
Our language customization gives you control over your news feed, while maintaining high factual accuracy. Click the **Tailor** button and move the sliders to change the following dimensions:

- **Complexity** *(regular, less complex)*: Makes the article more accessible so readers can understand the content easily and quickly. We replace difficult terms with simpler ones and add brief explanations for unfamiliar concepts.
- **Sentiment** *(more positive, regular, more negative)*: Changes the language of the article to be either more positive or negative. We highlight positive aspects and avoid negative terms to create a more positive sentiment, or emphasize potential downsides for a more negative sentiment.
- **Entertainment** *(regular, entertaining)*: Makes the article more entertaining by adding humor, puns, and clever word choices. We also rewrite some articles using a playful narrative style.
- **Political learning** *(liberal, regular, and conservative)*: Changes the political perspective of the article by using language and terms that are more commonly used by either conservatives or liberals. We also change the focus of the article to highlight themes and issues that matter most to each political group.

You can only change one dimension at a time (all other dimensions will be set back to "regular" if you change a second slider). You can change slider settings at any point in time, and the app will update the writing of all articles in real-time.

You will always see the same set of articles, no matter what your current customization looks like.


TAILORMADETIMES
Trending
Politics >
Regular
Carter Trump Legacy | 28min ago
Jimmy Carter's Legacy and Final Act of Civic Duty Contrasted with Trump's Return
Former President Jimmy Carter's passing at age 100, shortly after casting his final vote, highlights a stark contrast between his legacy of humanitarian efforts and Donald Trump's ongoing political journey.
Board Appointme... | 21min a...
Meta Expands Board with Additions from Sports, Business, and Tech Sectors
Regular
NIPPON STEEL
Trade Relations | 25min ago
Biden Blocks Nippon Steel's $15 Billion U.S.
→

## Some more information about the app

### Our sources

Each article is rewritten using advanced AI technology to deliver essential information quickly and clearly. The content is based on trending stories from the Associated Press (AP), ensuring professional journalism and high accuracy.

### Accurate reporting

While our AI improves the presentation, the underlying facts mentioned in the article remain unchanged from the AP sources, guaranteeing the same level of factual accuracy. We also have a human review of articles, just to be sure.

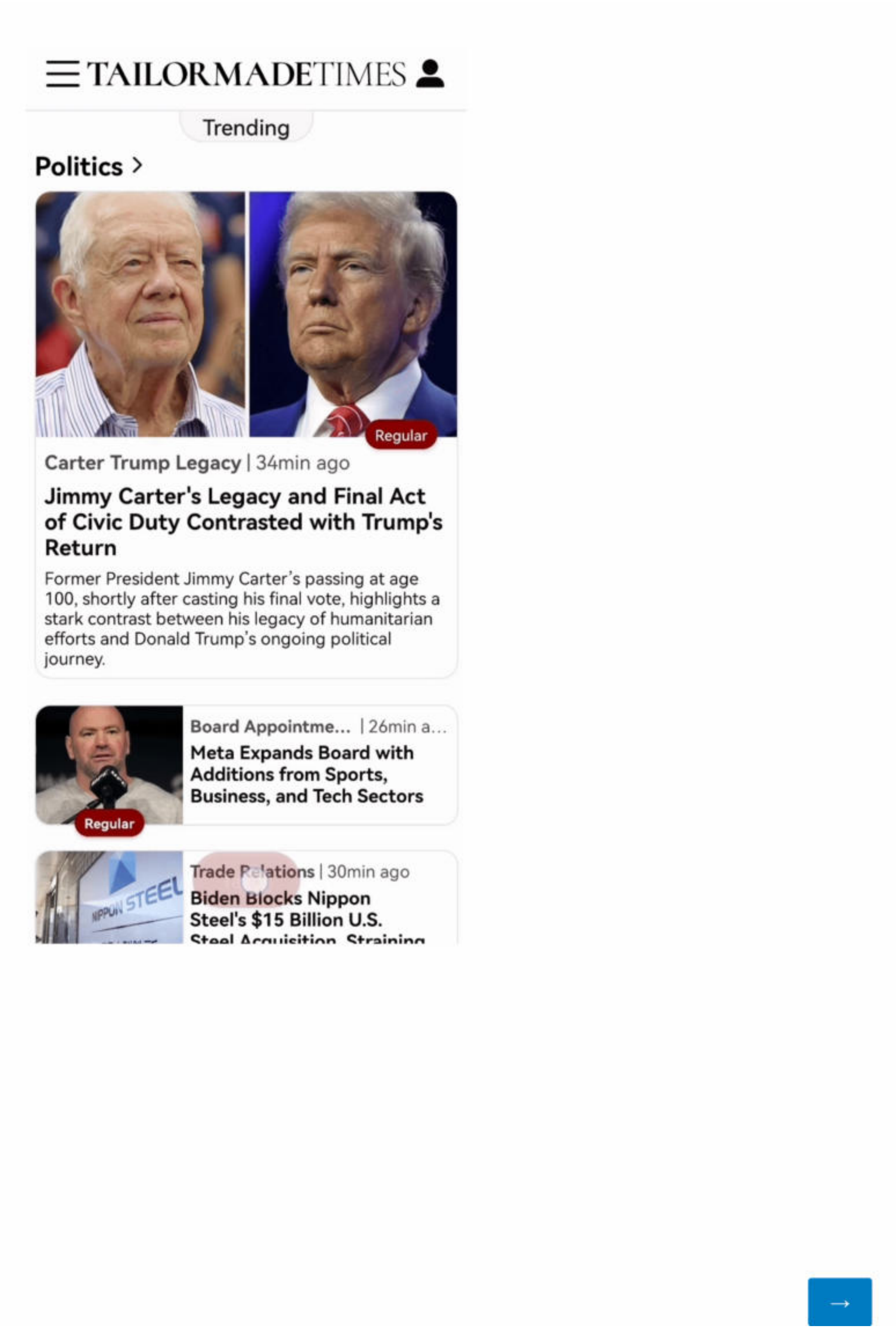
### Customization

Our language customization gives you control over your news feed, while maintaining high factual accuracy. Click the **Tailor** button and move the sliders to change the following dimensions:

- **Complexity** *(regular, less complex)*: Makes the article more accessible so readers can understand the content easily and quickly. We replace difficult terms with simpler ones and add brief explanations for unfamiliar concepts.
- **Sentiment** *(more positive, regular, more negative)*: Changes the language of the article to be either more positive or negative. We highlight positive aspects and avoid negative terms to create a more positive sentiment, or emphasize potential downsides for a more negative sentiment.
- **Entertainment** *(regular, entertaining)*: Makes the article more entertaining by adding humor, puns, and clever word choices. We also rewrite some articles using a playful narrative style.
- **Facts** *(regular, facts only): Makes the article less opinionated by focusing only on the facts.* Also avoids subjective assessments and judgements and uses more neutral language.

You can only change one dimension at a time (all other dimensions will be set back to "regular" if you change a second slider). You can change slider settings at any point in time, and the app will update the writing of all articles in real-time.

You will always see the same set of articles, no matter what your current customization looks like.


TAILORMADETIMES
Trending
Politics >
Regular
Carter Trump Legacy | 34min ago
Jimmy Carter's Legacy and Final Act of Civic Duty Contrasted with Trump's Return
Former President Jimmy Carter's passing at age 100, shortly after casting his final vote, highlights a stark contrast between his legacy of humanitarian efforts and Donald Trump's ongoing political journey.
Board Appointme... | 26min a...
Meta Expands Board with Additions from Sports, Business, and Tech Sectors
Regular
Trade Relations | 30min ago
Biden Blocks Nippon Steel's $15 Billion U.S.
NIPPON STEEL
→

How many dimensions can you tailor at a time?

- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4

What happens when you customize articles using the tailor button?

- ◯ I see a rewritten version of the same articles with the same underlying facts
- ◯ I see a rewritten version of the same articles with different facts
- ◯ I see different articles

→

*On the previous page, you provided an incorrect response to the following question:*

**How many dimensions can you tailor at a time?**

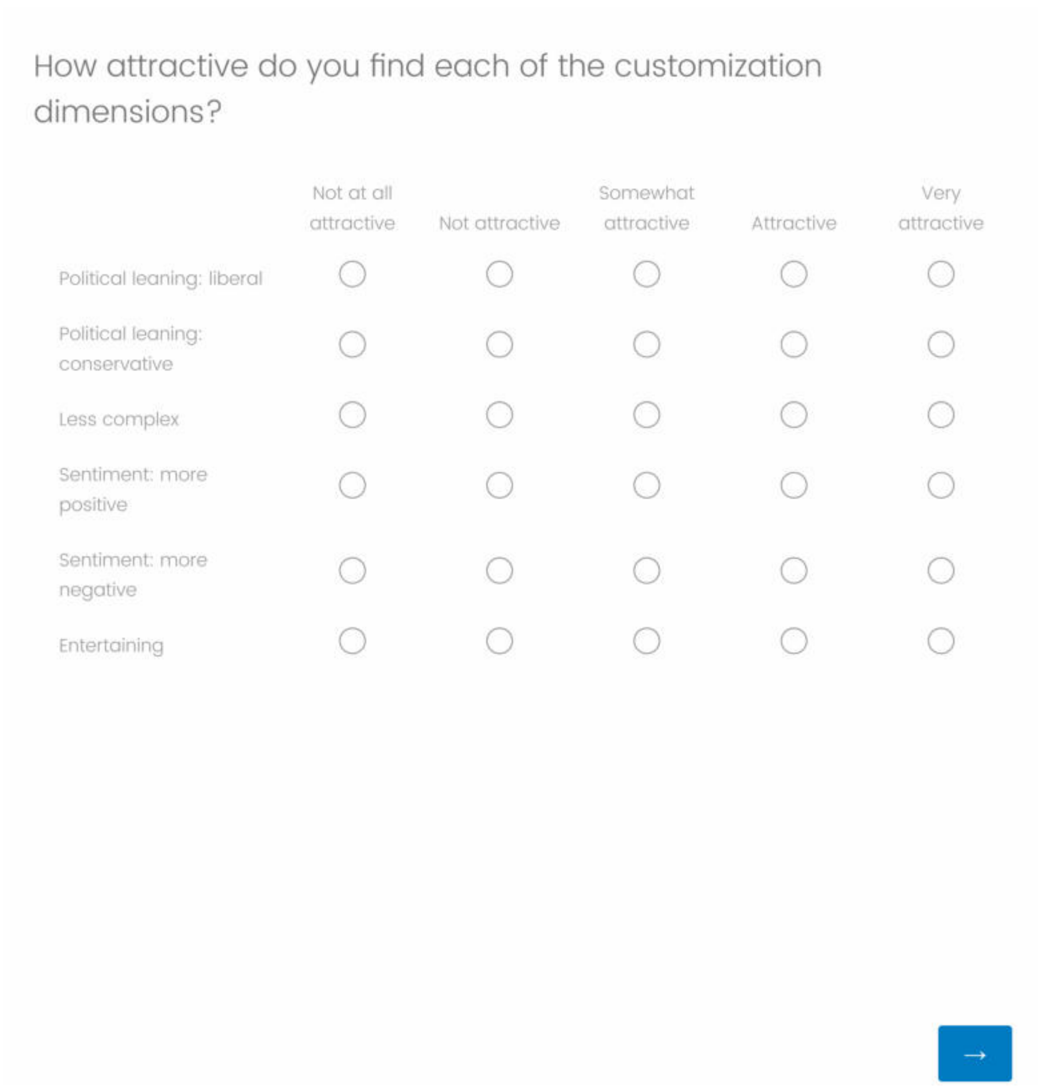
The correct answer is: "You can tailor **one dimension at a time**."

*On the previous page, you provided an incorrect response to the following question:*

**What happens when you customize articles using the tailor button?**

The correct answer is: "I see a **rewritten version** of the same articles **with the same underlying facts**."

→



How attractive do you find each of the customization dimensions?

| | Not at all attractive | Not attractive | Somewhat attractive | Attractive | Very attractive |
|---|---|---|---|---|---|
| Political leaning: liberal | ○ | ○ | ○ | ○ | ○ |
| Political leaning: conservative | ○ | ○ | ○ | ○ | ○ |
| Less complex | ○ | ○ | ○ | ○ | ○ |
| Sentiment: more positive | ○ | ○ | ○ | ○ | ○ |
| Sentiment: more negative | ○ | ○ | ○ | ○ | ○ |
| Entertaining | ○ | ○ | ○ | ○ | ○ |

→

How attractive do you find each of the customization dimensions?

| | Not at all attractive | Not attractive | Somewhat attractive | Attractive | Very attractive |
|---|---|---|---|---|---|
| Facts only | | | | | |
| Less complex | | | | | |
| Sentiment: more positive | | | | | |
| Sentiment: more negative | | | | | |
| Entertaining | | | | | |

→

## App use incentives: Earn $1 every day you use the app

Starting tomorrow and until February 6, you will receive an **additional $1 for each day that you read 3 or more articles** on our app.

Additionally, you will get an **additional $2 bonus** if you keep the app on your phone with push notifications for TailorMadeTimes **enabled** for the next three weeks (21 days). As proof of keeping the app and push notifications, you will have to open the article featured in the push notification after three weeks—we will send you a reminder the day before.

You will receive invitations for future surveys during this time.

Over the next three weeks, do you plan to keep push notifications for TailorMadeTimes enabled?

- ○ Yes, I plan to keep push notifications for TailorMadeTimes enabled.
- ○ No, I will not keep push notifications for TailorMadeTimes enabled.

→

## Screen time on your phone

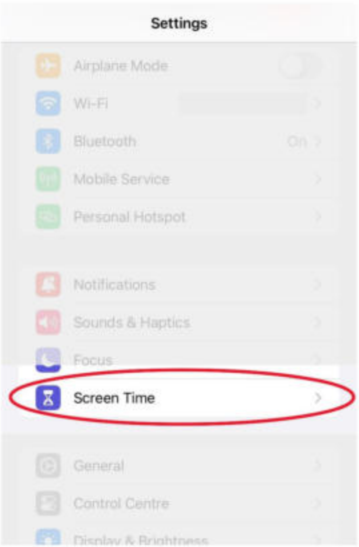
In 10 days, we will invite you for a follow-up survey where you can earn an **additional** bonus payment of $1 if you share a screenshot of your screen time with us.

→

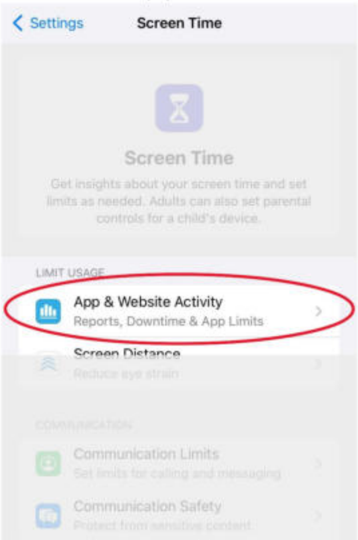
## Enabling screen time on your iPhone

In 10 days, we will invite you for a follow-up survey where you can earn an **additional** bonus payment of $1 if you share a screenshot of your screen time with us. To be able to do so, turn on screen time on your iPhone now:

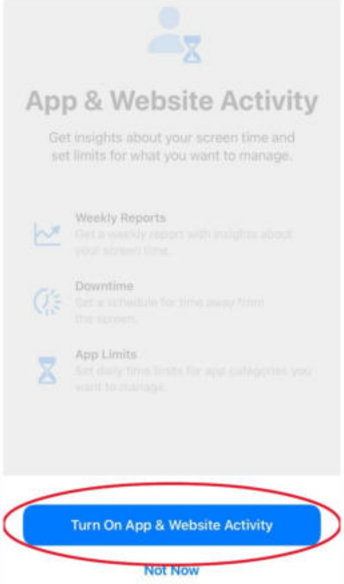
1. Go to settings
2. Click on "Screen Time"




3. Click on "App & Website Activity"



4. Click on "Turn on App & Website Activity"




(If it is already on, skip Step 4.)

Have you turned on screen time on your phone?

- ◯ Yes
- ◯ No

→

As part of this study, we have offered you many opportunities to earn additional bonuses over the next three weeks. Here is an overview to make sure you do not miss out on any of them.

**App use**: Starting tomorrow and until February 6, you can earn an additional **$1 for each day that you read three or more articles** on our app.

**Push notifications**: Earn **$2 for keeping push notifications enabled** for the TailorMadeTimes app for the next three weeks.

**Screen time**: Earn **$1 if you share a screenshot of your app usage** with us in a future survey.

If you seize all of the above bonus opportunities, you can earn **up to $13** over the next three weeks in addition to the $2 bonus that you will receive for having downloaded our app and enabled push. We will also invite you to future surveys over the next three weeks.

→

As part of this study, we have offered you many opportunities to earn additional bonuses over the next three weeks. Here is an overview to make sure you do not miss out on any of them.

**App use**: Starting tomorrow and until February 6, you can earn an additional **$1 for each day that you read three or more articles** on our app.

**Push notifications**: Earn **$2 for keeping push notifications enabled** for the TailorMadeTimes app for the next three weeks.

**Screen time**: Earn **$1 if you keep screen time enabled** on your phone and share a screenshot with us in a future survey.

If you seize all of the above bonus opportunities, you can earn **up to $13** over the next three weeks in addition to the $2 bonus that you will receive for having downloaded our app and enabled push. We will also invite you to future surveys over the next three weeks.

→

Thank you for participating!

→

### D.3.2 Midline Survey

This study is conducted by researchers from the Frankfurt School of Finance & Management, NHH Norwegian School of Economics, University of Cologne, and MIT. You must be at least 18 years of age to participate in this study. If you do not fulfill these requirements, please do not continue any further.

The goal of this study is to assess an AI-tailored news app, focusing on user experience and content relevance.

Your participation will help understanding people's user experience and improve the app. We do not anticipate any risks or discomforts from participating in this study.

Your responses to this survey will be anonymous; we will not collect your name or any other personal identifiers. We will only collect your Prolific ID to track your participation in this study, administer potential bonus payments, and invite you to potential future follow-up studies.

You are not allowed to participate in this study more than once. If you experience a technical error or problem, do not try to restart or retake the study. Rather, send us an email with a description of your problem and we will get back to you.

If you have any questions regarding this study, please email roth@wiso.uni-koeln.de

If you feel you have been treated unfairly, or you have questions regarding your rights as a research subject, you may contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, M.I.T., Room E25-143b, 77 Massachusetts Ave, Cambridge, MA 02139, phone 1-617-253-6787.

**Do you consent to participate in this study?**

- ○ I consent
- ○ I do not consent and wish to exit the study

→

We would like you to rate your feelings toward some different groups and individuals with whom you may agree or disagree politically. Ratings are made on what we call a 'feeling thermometer.' The scale ranges from 0 to 100. When you rate an individual or group, a score of 0 degrees means you feel as cold and unfavorable as possible towards them. A score of 100 degrees means you feel as warm and favorable as possible. You would rate them at 50 degrees if you do not feel particularly warm or cold toward them.

**Please rate how you feel about the following groups or individuals.**

Cold and unfavorable — Warm and favorable

0 10 20 30 40 50 60 70 80 90 100

**Republicans**

**Democrats**

→

Imagine that you had to split 100 dollars between two other people, Person A and Person B. Person A is a randomly chosen person among those who voted for the Democrat party in the last presidential election, while Person B is a randomly chosen person among those who voted for the Republican party in the last presidential election.

How much money would you like to give to Person A and how much money would you like to give to Person B?

| | |
|---|---|
| **Person A** (a Democrat) | 0 |
| **Person B** (a Republican) | 0 |
| Total | 0 |

→

Tell me whether you have a very favorable, somewhat favorable, somewhat unfavorable or very unfavorable opinion of **Donald Trump**.

- ○ Very favorable
- ○ Somewhat favorable
- ○ Somewhat unfavorable
- ○ Very unfavorable

Tell me whether you have a very favorable, somewhat favorable, somewhat unfavorable or very unfavorable opinion of **Joe Biden**.

- ○ Very favorable
- ○ Somewhat favorable
- ○ Somewhat unfavorable
- ○ Very unfavorable

Tell me whether you have a very favorable, somewhat favorable, somewhat unfavorable or very unfavorable opinion of **Kamala Harris**.

- ○ Very favorable
- ○ Somewhat favorable
- ○ Somewhat unfavorable
- ○ Very unfavorable

→

Overall, do you think that free trade agreements between the US and other countries have been a good thing or a bad thing for the United States?

- ○ Good thing
- ○ Bad thing

Overall, would you say that blacks or whites are treated more fairly in dealing with the police?

- ○ Blacks
- ○ Whites

Do you think that employers firing men who have been accused of sexual harassment or assault before finding out all the facts is a major or a minor problem?

- ○ Major problem
- ○ Minor problem

On the whole, do you think immigration is a good thing or a bad thing for this country today?

- ◯ Good thing
- ◯ Bad thing

In general, do you feel that the laws covering the sale of firearms should be made less strict, more strict, or kept as they are now?

- ◯ Less strict
- ◯ Kept as they are
- ◯ More strict

→

Over the current presidential turn, how much higher or lower do you expect the average **inflation rate** to be under President Trump compared to what it would have been had Harris won the presidency?

- ○ More than 10 percentage points higher under Trump
- ○ Between 5 and 10 percentage points higher under Trump
- ○ Between 2 and 5 percentage points higher under Trump
- ○ Between 0 and 2 percentage points higher under Trump
- ○ Between 0 and 2 percentage points lower under Trump
- ○ Between 2 and 5 percentage points lower under Trump
- ○ Between 5 and 10 percentage points lower under Trump
- ○ More than 10 percentage points lower under Trump

Over the current presidential turn, how much higher or lower do you expect the average **crime rate** to be under President Trump compared to what it would have been had Harris won the presidency?

- ○ More than 10 percentage points higher under Trump
- ○ Between 5 and 10 percentage points higher under Trump
- ○ Between 2 and 5 percentage points higher under Trump
- ○ Between 0 and 2 percentage points higher under Trump
- ○ Between 0 and 2 percentage points lower under Trump
- ○ Between 2 and 5 percentage points lower under Trump
- ○ Between 5 and 10 percentage points lower under Trump
- ○ More than 10 percentage points lower under Trump

Over the current presidential turn, how much higher or lower do you expect the average **unemployment rate** to be under President Trump compared to what it would have been had Harris won the presidency?

- More than 10 percentage points higher under Trump
- Between 5 and 10 percentage points higher under Trump
- Between 2 and 5 percentage points higher under Trump
- Between 0 and 2 percentage points higher under Trump
- Between 0 and 2 percentage points lower under Trump
- Between 2 and 5 percentage points lower under Trump
- Between 5 and 10 percentage points lower under Trump
- More than 10 percentage points lower under Trump

→

Powered by Qualtrics

Over the past two weeks, how often have you been bothered by any of the following problems?

| | Not at all | On several days | Nearly half of the days | More than half of the days | Every day |
|---|---|---|---|---|---|
| Little interest or pleasure in doing things | ○ | ○ | ○ | ○ | ○ |
| Feeling down, depressed or hopeless | ○ | ○ | ○ | ○ | ○ |

Overall, how happy have you felt over the past few days?

- ○ Very happy
- ○ Happy
- ○ Neutral
- ○ Unhappy
- ○ Very unhappy

→

## News quiz

On the next pages, we will ask you **15 factual questions** about the news coverage of the past few days. Please answer according to the best of your current knowledge and do not consult any other sources while answering these questions.

You will get a bonus payment based on your answers:

- **20 cents** for each question answered correctly.
- You need to answer each question within **20 seconds** to qualify for the bonus

→

What action did Canada take in response to the tariffs imposed by President Trump?

- ○ They did not offer an immediate response and instead attempted to bargain directly with President Trump.
- ○ They banned imports of select American dairy products.
- ○ They threatened to abandon NAFTA unless the tariffs were rescinded.
- ○ They lodged a complaint with the World Trade Organization.
- ○ They imposed 25% retaliatory tariffs on $155 billion worth of American products.

→

What concern has been raised by OpenAI and U.S. officials regarding the Chinese tech company DeepSeek?

- ○ DeepSeek openly acknowledged using OpenAI's models without permission.
- ○ DeepSeek may have used ChatGPT technology to develop its AI model.
- ○ DeepSeek has filed a lawsuit against OpenAI for patent infringement.
- ○ DeepSeek is accused of hacking into OpenAI's systems to steal data.
- ○ DeepSeek has admitted to using OpenAI's source code for its AI.

→

What executive order from President Trump has led to the removal of certain public data from government websites?

- ◯ An order to promote the publication of taxpayer-funded datasets
- ◯ An order to strip references to "gender ideology" from federal materials
- ◯ An order to increase the transparency of climate change data
- ◯ An order to remove all references to healthcare statistics
- ◯ An order to enhance vaccine-related information on federal websites

→

What was the primary stated purpose of President Trump's announced tariffs on Canadian and Mexican imports?

- ◯ To retaliate against trade imbalances with Canada and Mexico
- ◯ To increase government revenue and reduce the national deficit
- ◯ To promote domestic manufacturing and reduce reliance on foreign goods
- ◯ To curb illegal immigration and the smuggling of chemicals used in fentanyl production
- ◯ To stabilize oil prices by limiting imports from Canada and Mexico

→

Why was there recently a dispute between President Trump and Colombia?

- ○ Trump threatened a 25% tariff on Colombian exports unless Colombia accepted U.S. flights carrying deported immigrants.
- ○ Trump warned of visa restrictions on Colombian officials unless Colombia allowed U.S. deported immigrant flights.
- ○ Trump insisted on increased U.S.-Colombia military cooperation, threatening aid cuts unless deported immigrants were repatriated.
- ○ Trump threatened to impose tariffs on Colombian crude oil imports unless Colombia revised its overall trade policies.
- ○ Trump demanded that Colombia boost its flower exports for Valentine's Day, or else face new trade tariffs.

→

What is a key reason for U.S. Secretary of State Marco Rubio's visit to Panama?

- ○ To finalize a trade agreement between the U.S. and Panama.
- ○ To promote tourism from the U.S. to Panama.
- ○ To discuss Panama's purchase of U.S. military equipment.
- ○ To oversee the construction of a new canal in Panama.
- ○ To negotiate the U.S. regaining control of the Panama Canal.

→

What controversial proposal did President Trump make in relation to the Middle East conflict?

- ○ He proposed mediating direct negotiations between Israel and Hamas to secure a lasting ceasefire.
- ○ He stated that a two-state solution would never work in practice and should no longer be discussed.
- ○ He proposed bolstering the U.S. presence in the Middle East to protect Israel against its enemies.
- ○ He proposed threatening Saudi Arabia with tariffs if it did not support normalization of diplomatic efforts with Israel.
- ○ He proposed that the U.S. seize control of Gaza, potentially by force.

→

What roadblock did House Republicans encounter while crafting a bill to extend and expand the tax cuts they passed in President Donald Trump's first term?

- ○ Gathering support from Elon Musk, who wanted the tax cuts to be even more aggressive
- ○ Figuring out how to pay for the multitrillion-dollar cost of extending and expanding the tax cuts
- ○ Determining how to reconcile Medicaid spending cuts with other budget priorities
- ○ Convincing Democrats to support the tax bill
- ○ Resolving disagreements over the allocation of funds for border security

→

What major initiative did India's government recently announce?

- ○ Increased tariffs on US exports in response to tariff threats by President Trump.
- ○ A comprehensive plan to boost digital learning and invest in artificial intelligence.
- ○ A new budget focusing on middle-class tax relief, agriculture, manufacturing, and clean energy to spur economic growth.
- ○ A ban on imports of Chinese electric vehicles.
- ○ An initiative to improve relations with the US and the newly elected Trump administration.

→

How did the stock market react to President Trump's tariff threats on imports from Canada, Mexico, and China?

- ○ Stocks showed a mixed performance, with some sectors—such as technology and automotive—posting gains while others declined.
- ○ Stocks rallied as investors anticipated that higher tariffs would boost domestic manufacturing.
- ○ Stocks across multiple sectors fell sharply as investors grew increasingly concerned about trade disruptions and rising production costs.
- ○ The overall market remained largely unchanged, with traders expecting the tariff threats to be only temporary.
- ○ The overall market registered an increase, although oil companies experienced significant drops amid the trade tensions.

→

Why was the U.S. Postal Service recently featured in the news?

- It introduced an expedited international shipping service to meet increased demand from overseas retailers.
- It launched a new digital tracking system for international packages to improve delivery times.
- It suspended parcel deliveries from China and Hong Kong following Trump's executive order that imposed a 10% tariff on Chinese goods.
- It reversed a longstanding policy that allowed small-value packages to enter the U.S. tax-free in order to enforce stricter tariff regulations.
- It partnered with private carriers to expand its service into additional global markets.

→

What controversial actions has Elon Musk taken as part of his role in overseeing the Department of Government Efficiency?

- ○ Making all forms of affirmative action illegal for all government agencies.
- ○ Implementing large pay cuts for government officials.
- ○ Requiring all federal agencies to cut their staff by 20% by the end of the year.
- ○ Moved to fire or sideline career government officials and gained access to sensitive taxpayer databases.
- ○ Implementing transparency measures to give the public more oversight of government spending and operations.

→

What did President Donald Trump claim his administration has been doing in relation to the war in Ukraine?

- ◯ Establishing a no-fly zone over Ukraine to protect civilians
- ◯ Funding humanitarian aid exclusively for Ukrainian refugees
- ◯ Imposing new sanctions on Russia to pressure them into ending the war
- ◯ Engaging in "very serious" discussions with Russia to address the conflict
- ◯ Collaborating with NATO to increase military support for Ukraine

→

Why were the Marines recently in the news?

- ○ They bolstered border security as part of President Trump's efforts by installing extra concertina wire along the border wall.
- ○ A senior general warned about insufficient equipment to handle multiple simultaneous threats facing the U.S.
- ○ They conducted tactical exercises close to Greenland.
- ○ They participated in a joint training exercise with local law enforcement to improve border patrol tactics.
- ○ They scrapped all Diversity, Equity, and Inclusion (DEI) training initiatives.

→

What controversial executive order was recently blocked in court?

- ◯ Trump's order to end birthright citizenship for children born in the U.S. to undocumented immigrants.
- ◯ Trump's order to reallocate military funds for enhanced border security.
- ◯ Trump's order to reduce federal funding for the Environmental Protection Agency.
- ◯ Trump's order to dismantle the Deferred Action for Childhood Arrivals (DACA) program.
- ◯ Trump's order to require mandatory voter ID for federal elections.

→

**Your experience with the app**

A few days ago, you were offered early access to our news app, *TailorMadeTimes*. We would now like to learn more about your experience using the app over the past few days.

TAILORMADETIMES

→

How satisfied are you with the news that you have been reading over the past week?

- ◯ Very satisfied
- ◯ Somewhat satisfied
- ◯ Neutral
- ◯ Somewhat dissatisfied
- ◯ Very dissatisfied

→

How did access to the app affect your time spent consuming news from **other news sources** over the past few days?

- ◯ Spent **much less** time consuming news from other news sources than usual
- ◯ Spent **slightly less** time consuming news from other news sources than usual
- ◯ Spent **about the same** time consuming news from other news sources as usual
- ◯ Spent **slightly more** time consuming news from other news sources than usual
- ◯ Spent **much more** time consuming news from other news sources than usual

How did access to the app affect your **total time** consuming news over the past few days?

- ◯ Spent **much less** time consuming news than usual
- ◯ Spent **slightly less** time consuming news than usual
- ◯ Spent **about the same** time consuming news as usual
- ◯ Spent **slightly more** time consuming news than usual
- ◯ Spent **much more** time consuming news than usual

→

Over the past few days, relative to what is typical for you, would you say you spent more or less of your free time (i.e. excluding work time) on...?

| | A lot less | Somewhat less | Same | Somewhat more | A lot more |
|---|---|---|---|---|---|
| using social media apps? | | | | | |
| online (on your computer, tablet, smartphone, etc.) for things other than social media? | | | | | |
| watching TV or movies by yourself? | | | | | |
| on non-screen activities (e.g. cooking, reading books, exercising– anything without an electronic screen in front of you) by yourself? | | | | | |
| doing anything with friends and family (in person)? | | | | | |

→

How complex or simple do you find the articles in the TailorMadeTimes app?

- ○ Very simple
- ○ Simple
- ○ Neither simple nor complex
- ○ Complex
- ○ Very complex

How would you rate the political bias of the articles in the TailorMadeTimes app?

- ○ Very right-wing biased
- ○ Somewhat right-wing biased
- ○ Not biased
- ○ Somewhat left-wing biased
- ○ Very left-wing biased

How would you rate the quality of articles in the TailorMadeTimes app?

- ○ Very high quality
- ○ High quality
- ○ Medium quality
- ○ Low quality
- ○ Very low quality

How would you characterize the articles on TailorMadeTimes on a continuum from opinionated to fact-driven?

- Very opinionated
- Somewhat ppinionated
- Balanced/Neutral
- Somewhat fact-friven
- Very fact-driven

How entertaining do you find the articles in the TailorMadeTimes app?

- Very entertaining
- Entertaining
- Somewhat entertaining
- Not entertaining
- Not entertaining at all

How accurate are the articles in the TailorMadeTimes app?

- Very accurate
- Accurate
- Somewhat accurate
- Inaccurate
- Very inaccurate

How trustworthy do you find the articles in the TailorMadeTimes app?

- Very trustworthy
- Trustworthy
- Somewhat trustworthy
- Not trustworthy
- Not trustworthy at all

How would you rate the sentiment of the articles in the TailorMadeTimes app?

- Very negative sentiment
- Somewhat negative sentiment
- Neutral sentiment
- Somewhat positive sentiment
- Very positive sentiment

→

How much do you trust AI technology, such as language models, in general?

- ◯ A great deal
- ◯ A lot
- ◯ A moderate amount
- ◯ A little
- ◯ None at all

How much do you trust AI to accurately and impartially tailor news about politics and the economy?

- ◯ A great deal
- ◯ A lot
- ◯ A moderate amount
- ◯ A little
- ◯ None at all

→

To what extent do you agree with the following statement: *AI technology has the potential to make the news more relevant for me.*

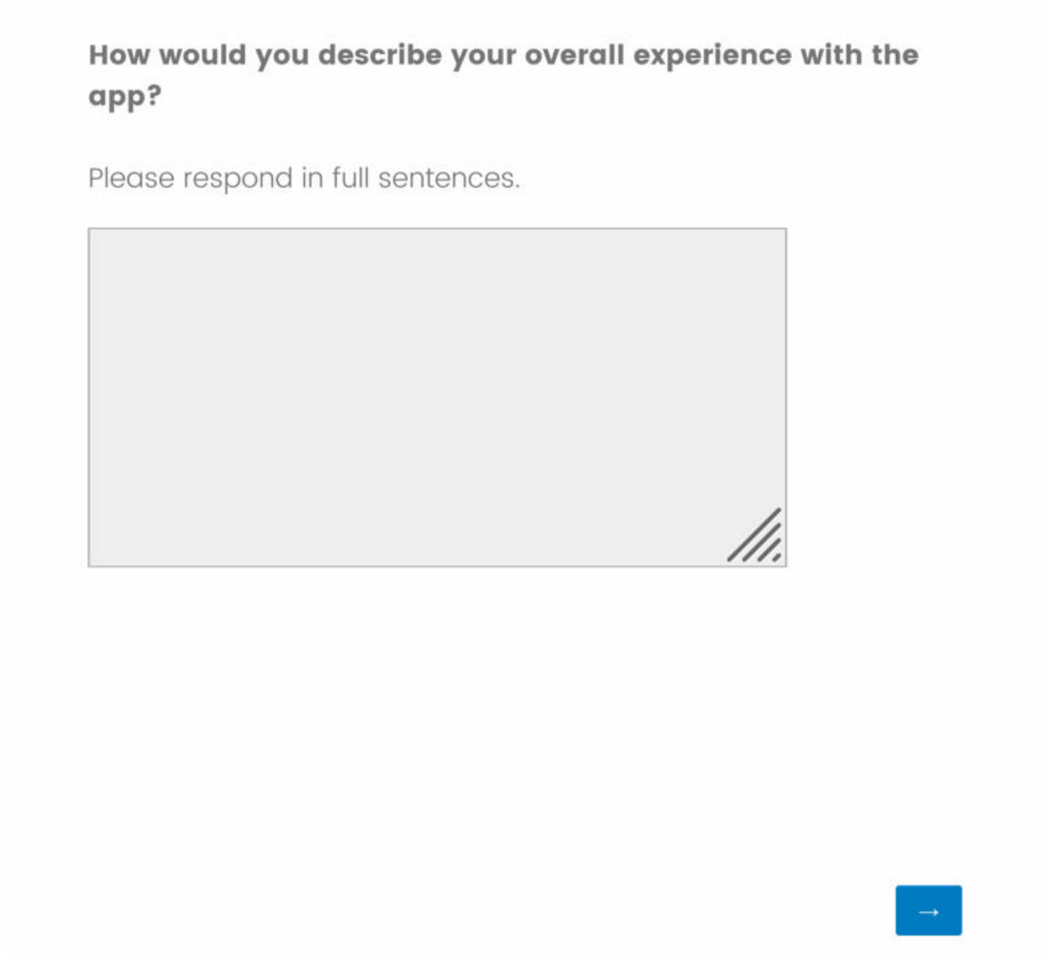
- ◯ Strongly agree
- ◯ Somewhat agree
- ◯ Neither agree not disagree
- ◯ Somewhat disagree
- ◯ Strongly disagree

→

How often did you use the app?

- ◯ Never
- ◯ Once
- ◯ A few times
- ◯ Multiple times, but not daily
- ◯ Daily
- ◯ Multiple times a day

→


How would you describe your overall experience with the app?
Please respond in full sentences.
→

What do you think about the **length of the articles** in the app?

- ○ Too short
- ○ Somewhat short
- ○ About the right length
- ○ Somewhat long
- ○ Too long

What do you think about the **update frequency of the articles** in the app?

- ○ Updated too frequently
- ○ Updated somewhat too frequently
- ○ Updated about at the right frequency
- ○ Updated somewhat too infrequently
- ○ Updated too infrequently

What do you think about the **number of articles** in the app?

- ○ Too few articles
- ○ Somewhat too few articles
- ○ About the right number of articles
- ○ Somewhat too many articles
- ○ Too many articles

What do you think about the **number of push notifications** by the app?

- ○ Too few notifications
- ○ Somewhat too few notifications
- ○ About the right number of notifications
- ○ Somewhat too many notifications
- ○ Too many notifications

How satisfied were you with the **selection of the articles** included in the app?

- ○ Very satisfied
- ○ Somewhat satisfied
- ○ Neutral
- ○ Somewhat dissatisfied
- ○ Very dissatisfied

→

How satisfied were you with your **overall experience** with the app?

- ◯ Very satisfied
- ◯ Somewhat satisfied
- ◯ Neutral
- ◯ Somewhat dissatisfied
- ◯ Very dissatisfied

→

Did you encounter any technical issues with the app?

- ◯ Yes
- ◯ No

→

Compared to the regular article, do you think the customized versions have higher, lower, or the same factual accuracy?

- ◯ Much higher
- ◯ Higher
- ◯ Same
- ◯ Lower
- ◯ Much lower

How would you rate the overall article quality for each of the following rewrites compared to the regular article?

| | Much lower quality than regular | Lower quality than regular | About the same quality as regular | Higher quality than regular | Much higher quality than regular |
|---|---|---|---|---|---|
| Conservative | ◯ | ◯ | ◯ | ◯ | ◯ |
| Less complex | ◯ | ◯ | ◯ | ◯ | ◯ |
| Positive | ◯ | ◯ | ◯ | ◯ | ◯ |
| Liberal | ◯ | ◯ | ◯ | ◯ | ◯ |
| Entertaining | ◯ | ◯ | ◯ | ◯ | ◯ |
| Negative | ◯ | ◯ | ◯ | ◯ | ◯ |

→

Compared to the regular article, do you think the customized versions have higher, lower, or the same factual accuracy?

- ◯ Much higher
- ◯ Higher
- ◯ Same
- ◯ Lower
- ◯ Much lower

How would you rate the overall article quality for each of the following rewrites compared to the regular article?

| | Much lower quality than regular | Lower quality than regular | About the same quality as regular | Higher quality than regular | Much higher quality than regular |
|---|---|---|---|---|---|
| Facts only | ◯ | ◯ | ◯ | ◯ | ◯ |
| Entertaining | ◯ | ◯ | ◯ | ◯ | ◯ |
| Positive | ◯ | ◯ | ◯ | ◯ | ◯ |
| Negative | ◯ | ◯ | ◯ | ◯ | ◯ |
| Less complex | ◯ | ◯ | ◯ | ◯ | ◯ |

→

How attractive do you find the different customization options?

| | Not at all attractive | Not attractive | Somewhat attractive | Attractive | Very attractive |
|---|---|---|---|---|---|
| **Sentiment** (more positive, regular, more negative) | ○ | ○ | ○ | ○ | ○ |
| **Political leaning** (liberal, regular, conservative) | ○ | ○ | ○ | ○ | ○ |
| **Complexity** (regular, less complex) | ○ | ○ | ○ | ○ | ○ |
| **Entertainment** (regular, entertaining) | ○ | ○ | ○ | ○ | ○ |

→

How attractive do you find the different customization options?

| | Not at all attractive | Not attractive | Somewhat attractive | Attractive | Very attractive |
|---|---|---|---|---|---|
| **Complexity** (regular, less complex) | ○ | ○ | ○ | ○ | ○ |
| **Sentiment** (more positive, regular, more negative) | ○ | ○ | ○ | ○ | ○ |
| **Entertainment** (regular, entertaining) | ○ | ○ | ○ | ○ | ○ |
| **Facts** (regular, facts only) | ○ | ○ | ○ | ○ | ○ |

→

## Bonus part: Screen Time

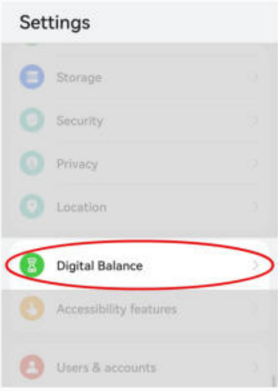
In this part, you can get a $2 bonus payment. Your base payment for the survey does not depend on the answers you give in this part.

Screen Time is a feature on phones that provides a summary of how much time you spend on different apps.

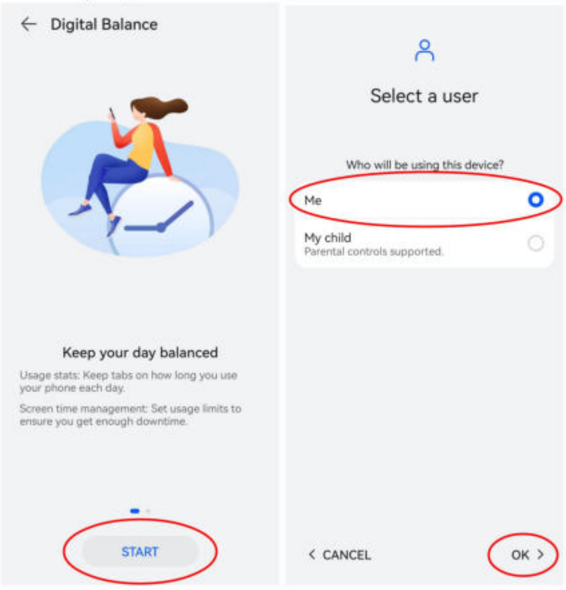
You can now earn an **additional bonus of $2** for sharing a screenshot of the Screen Time page from your phone.

Here are step-by-step instructions on how to do this:

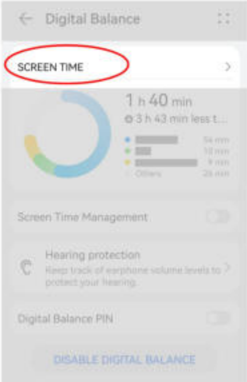
1. Go to settings.
2. Click on "Digital Balance"




3. If it is not activated yet, click on "Start", then select "Me" and click "OK". (If it is already activated, skip this step and continue with Step 4.)

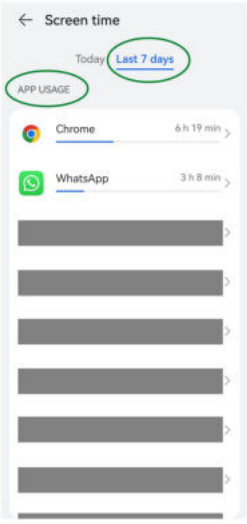


4. Click on "SCREEN TIME".




5. Click on "Last 7 days" at the top and then on "Show more" at the bottom.


6. Scroll down to the list of apps. Now take a screenshot. It should look like this:
(The date range should be "Last 7 days" and it should show "APP USAGE" with a list of apps and time.)



7. Upload the screenshots below.

**Please upload the screenshot of your app usage from your phone below.**

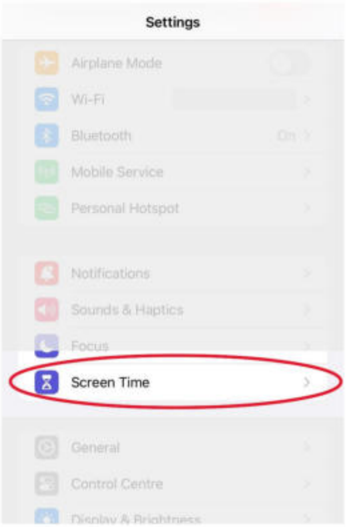
Drop files or click here to upload

→

## Bonus part: Screen Time

In this part, you can get a $2 bonus payment. Your base payment for the survey does not depend on the answers you give in this part.

Screen Time is a feature on iPhones that provides a summary of how much time you spend on different apps.
You can find Screen Time by going to **Settings** and scrolling down until you see **Screen Time**:




**Do you currently have Screen Time activated on your iPhone?**

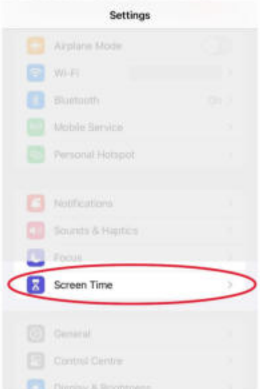
- ◯ Yes
- ◯ No

→

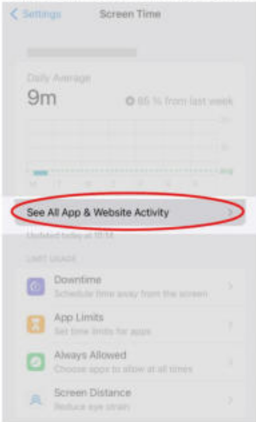
You can now earn an **additional bonus of $2** for sharing a screenshot of the Screen Time page from your iPhone.

Here are step-by-step instructions on how to do this:

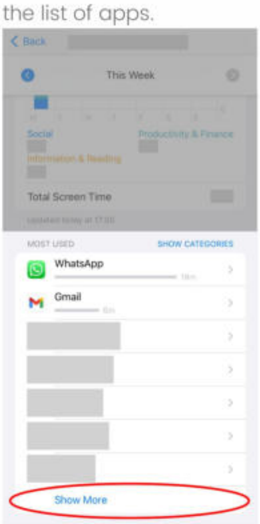
1. Go to settings.
2. Click on "Screen Time"



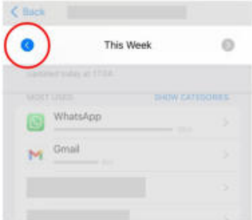

3. Click on "See All App & Website Activity".



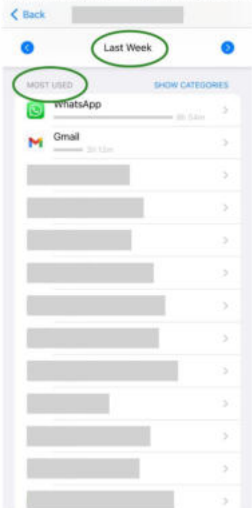

4. Scroll down to "MOST USED" and click on "Show More" below the list of apps.



5. Click the arrow at top left to go to previous week.




6. Now take a screenshot. It should look like this:
(The date range should be "Last Week" and it should show "MOST USED" with a list of apps and time.)




7. Upload the screenshots below.

**Please upload the screenshot of your app usage from your phone below.**

Drop files or click here to upload

→


Thank you for participating!

If there are any remarks that you would like to make or clarifications that you would like to obtain, please do let us know by writing them into the field below.

→

### D.3.3 Endline Survey

This study is conducted by researchers from the Frankfurt School of Finance & Management, NHH Norwegian School of Economics, University of Cologne, and MIT. You must be at least 18 years of age to participate in this study. If you do not fulfill these requirements, please do not continue any further.

The goal of this study is to assess an AI-tailored news app, focusing on user experience and content relevance.

Your participation will help understanding people's user experience and improve the app. We do not anticipate any risks or discomforts from participating in this study.

Your responses to this survey will be anonymous; we will not collect your name or any other personal identifiers. We will only collect your Prolific ID to track your participation in this study, administer potential bonus payments, and invite you to potential future follow-up studies.

You are not allowed to participate in this study more than once. If you experience a technical error or problem, do not try to restart or retake the study. Rather, send us an email with a description of your problem and we will get back to you.

If you have any questions regarding this study, please email roth@wiso.uni-koeln.de

If you feel you have been treated unfairly, or you have questions regarding your rights as a research subject, you may contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, M.I.T., Room E25-143b, 77 Massachusetts Ave, Cambridge, MA 02139, phone 1-617-253-6787.

**Do you consent to participate in this study?**

- ○ I consent
- ○ I do not consent and wish to exit the study

→

We would like you to rate your feelings toward some different groups and individuals with whom you may agree or disagree politically. Ratings are made on what we call a 'feeling thermometer.' The scale ranges from 0 to 100. When you rate an individual or group, a score of 0 degrees means you feel as cold and unfavorable as possible towards them. A score of 100 degrees means you feel as warm and favorable as possible. You would rate them at 50 degrees if you do not feel particularly warm or cold toward them.

**Please rate how you feel about the following groups or individuals.**

Cold and unfavorable Warm and favorable

0 10 20 30 40 50 60 70 80 90 100

**Democrats**

**Republicans**

→

Imagine that you had to split 100 dollars between two other people, Person A and Person B. Person A is a randomly chosen person among those who voted for the Democrat party in the last presidential election, while Person B is a randomly chosen person among those who voted for the Republican party in the last presidential election.

How much money would you like to give to Person A and how much money would you like to give to Person B?

| | |
|---|---|
| **Person A** (a Democrat) | 0 |
| **Person B** (a Republican) | 0 |
| Total | 0 |

→

Tell me whether you have a very favorable, somewhat favorable, somewhat unfavorable or very unfavorable opinion of **Donald Trump**.

- ○ Very favorable
- ○ Somewhat favorable
- ○ Somewhat unfavorable
- ○ Very unfavorable

Tell me whether you have a very favorable, somewhat favorable, somewhat unfavorable or very unfavorable opinion of **Joe Biden**.

- ○ Very favorable
- ○ Somewhat favorable
- ○ Somewhat unfavorable
- ○ Very unfavorable

Tell me whether you have a very favorable, somewhat favorable, somewhat unfavorable or very unfavorable opinion of **Kamala Harris**.

- ○ Very favorable
- ○ Somewhat favorable
- ○ Somewhat unfavorable
- ○ Very unfavorable

→

Overall, do you think that free trade agreements between the US and other countries have been a good thing or a bad thing for the United States?

- ○ Good thing
- ○ Bad thing

Overall, would you say that blacks or whites are treated more fairly in dealing with the police?

- ○ Blacks
- ○ Whites

Do you think that employers firing men who have been accused of sexual harassment or assault before finding out all the facts is a major or a minor problem?

- ○ Major problem
- ○ Minor problem

On the whole, do you think immigration is a good thing or a bad thing for this country today?

- ◯ Good thing
- ◯ Bad thing

In general, do you feel that the laws covering the sale of firearms should be made less strict, more strict, or kept as they are now?

- ◯ Less strict
- ◯ Kept as they are
- ◯ More strict

→

Over the current presidential turn, how much higher or lower do you expect the average **inflation rate** to be under President Trump compared to what it would have been had Harris won the presidency?

- More than 10 percentage points higher under Trump
- Between 5 and 10 percentage points higher under Trump
- Between 2 and 5 percentage points higher under Trump
- Between 0 and 2 percentage points higher under Trump
- Between 0 and 2 percentage points lower under Trump
- Between 2 and 5 percentage points lower under Trump
- Between 5 and 10 percentage points lower under Trump
- More than 10 percentage points lower under Trump

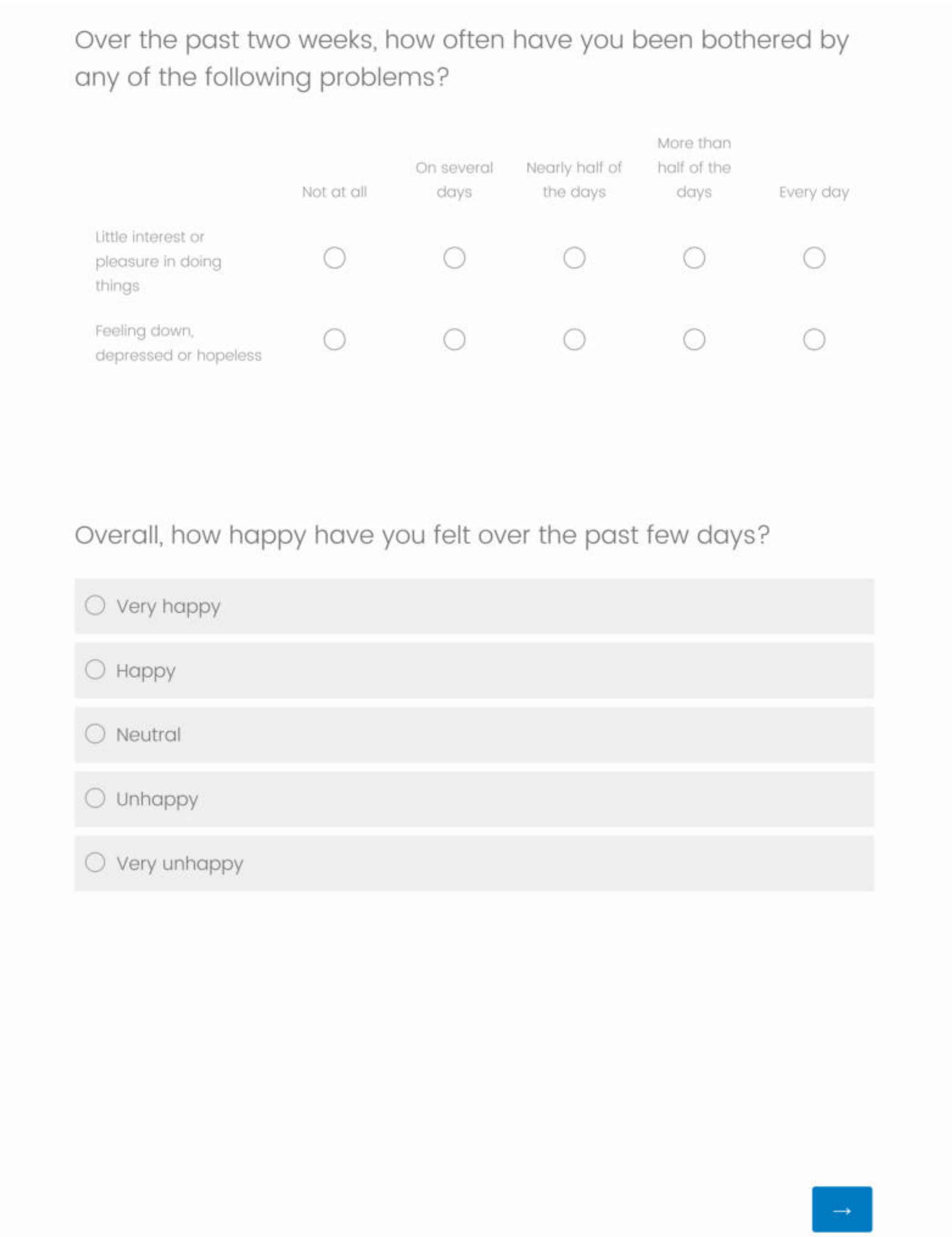
Over the current presidential turn, how much higher or lower do you expect the average **crime rate** to be under President Trump compared to what it would have been had Harris won the presidency?

- More than 10 percentage points higher under Trump
- Between 5 and 10 percentage points higher under Trump
- Between 2 and 5 percentage points higher under Trump
- Between 0 and 2 percentage points higher under Trump
- Between 0 and 2 percentage points lower under Trump
- Between 2 and 5 percentage points lower under Trump
- Between 5 and 10 percentage points lower under Trump
- More than 10 percentage points lower under Trump

Over the current presidential turn, how much higher or lower do you expect the average **unemployment rate** to be under President Trump compared to what it would have been had Harris won the presidency?

- ◯ More than 10 percentage points higher under Trump
- ◯ Between 5 and 10 percentage points higher under Trump
- ◯ Between 2 and 5 percentage points higher under Trump
- ◯ Between 0 and 2 percentage points higher under Trump
- ◯ Between 0 and 2 percentage points lower under Trump
- ◯ Between 2 and 5 percentage points lower under Trump
- ◯ Between 5 and 10 percentage points lower under Trump
- ◯ More than 10 percentage points lower under Trump

→



Over the past two weeks, how often have you been bothered by any of the following problems?

| | Not at all | On several days | Nearly half of the days | More than half of the days | Every day |
|---|---|---|---|---|---|
| Little interest or pleasure in doing things | ○ | ○ | ○ | ○ | ○ |
| Feeling down, depressed or hopeless | ○ | ○ | ○ | ○ | ○ |

Overall, how happy have you felt over the past few days?

- ○ Very happy
- ○ Happy
- ○ Neutral
- ○ Unhappy
- ○ Very unhappy

→

## News quiz

On the next pages, we will ask you **15 factual questions** about the news coverage of the past few days. Please answer according to the best of your current knowledge and do not consult any other sources while answering these questions.

You will get a bonus payment based on your answers:

- **20 cents** for each question answered correctly.
- You need to answer each question within **20 seconds** to qualify for the bonus

→

What hurdles is President Trump facing over his numerous executive orders targeting immigration?

- ○ Political pushback from Congress, with some lawmakers planning to override portions of the orders.
- ○ Mass resignations within Immigration and Customs Enforcement and other federal agencies.
- ○ Concerns from wealthy donors and conservative think tanks about potential labor shortages for essential seasonal work.
- ○ A lack of cooperation from host countries in managing the repatriation of deported individuals.
- ○ Legal challenges and logistical issues due to limited enforcement personnel and difficulties in coordinating among federal agencies.

→

What did the parent company of Truth Social, President Trump's social media platform, recently report?

- Plans to join a consortium to buy TikTok in the U.S.
- A strategic collaboration with the social media platform X and its owner Elon Musk.
- A $400.9 million profit and an increase in revenue.
- A $400.9 million loss and a drop in revenue.
- A massive boost in membership numbers after President Trump's inauguration.

→

What key proposal did Canadian Prime Minister Justin Trudeau make in response to President Trump's tariff threats?

- ○ He proposed immediate retaliatory tariffs on $155 billion CAD worth of U.S. goods.
- ○ He suggested a deal in which Canada would impose stricter border controls to appease President Trump.
- ○ He proposed a joint meeting with President Trump at Mar-a-Lago to discuss the issues over a round of golf.
- ○ He urged Canada to collaborate with the U.S. and expand global trade partnerships to mitigate the impact of potential tariffs.
- ○ He proposed coordinating countermeasures with the EU to exert stronger pressure on the U.S.

→

What action did the Trump administration recently take regarding wildfire mitigation funds?

- They added funds to enforce stricter oversight measures, ensuring that funds are properly allocated for wildfire mitigation efforts.
- They froze funding for wildfire mitigation projects under the Bipartisan Infrastructure Law and Inflation Reduction Act, sparking concerns about forest management and community safety.
- They encouraged a matching scheme in which states contribute their own funds to supplement federal wildfire mitigation support.
- They redirected existing wildfire mitigation funds to prioritize the hiring and training of seasonal wildland firefighters to enhance emergency readiness.
- They increased funding to private initiatives to bolster public-private collaboration on wildfire mitigation.

→

What new tariff initiative did President Trump recently announce?

- ○ A graduated tariff scale that increases duties based on a country's trade deficit with the U.S.
- ○ A measure to impose additional tariffs on imports from countries that fail to reciprocate U.S. market access
- ○ A "reciprocal tariffs" system, aiming to match foreign import tariffs.
- ○ A plan to implement temporary tariffs that automatically adjust based on economic indicators like inflation and trade balances.
- ○ A proposal to lower tariffs for allied nations in exchange for them purchasing U.S. military equipment.

→

Why has Illinois Governor J.B. Pritzker delayed the plan to prevent invasive carp from entering the Great Lakes?

- ◯ Sufficient natural barriers already preventing carp migration
- ◯ Lack of support from local environmental groups
- ◯ Rising costs of the project beyond the initial budget
- ◯ Concerns about the federal government's financial commitment
- ◯ Completion of the project ahead of schedule

→

What did a January survey reveal about American attitudes towards government spending?

- ○ Approximately two-thirds of Americans believe that the government overspends on social programs, including Social Security and education.
- ○ Only a small minority of Americans support increasing funding for programs like Medicare and Medicaid.
- ○ The survey found that most Americans favor significant cuts to government spending in areas such as healthcare and education.
- ○ About two-thirds of Americans believe that the government spends too little on Social Security and education.
- ○ The survey revealed that Americans are largely satisfied with current spending levels on social services.

→

What did President Donald Trump express interest in during an interview aired before the Super Bowl?

- ◯ Having Canada become the 51st state of the United States
- ◯ Ending all DEI initiatives on college campuses
- ◯ Imposing new tariffs on Canada
- ◯ Building a wall along the U.S.-Canada border
- ◯ Closing the border to Canada unless they stop all drug trafficking

→

What controversial proposal did President Trump make regarding Gaza?

- ○ He suggested annexing Gaza to Israel.
- ○ He suggested installing U.S. bases in Gaza to guarantee peace and security in the region.
- ○ He suggested a joint initiative to invest in the rebuilding of Gaza if Hamas agreed to release all hostages.
- ○ He proposed transforming Gaza into an international free trade zone to boost economic growth.
- ○ He suggested developing Gaza into a Mediterranean tourist destination.

→

What did Federal Reserve Chair Jerome Powell recently announce regarding the Fed's interest rate policies?

- ○ The Fed would consider raising its annual inflation target to 3%.
- ○ President Trump's tariff policies might allow the Fed to cut rates sooner than expected.
- ○ The Fed would end its current efforts to maintain a gender balance on its board.
- ○ The Fed would base its decisions solely on economic data and conditions, not on political pressure.
- ○ The Fed would take President Trump's calls for lower interest rates into account when deciding on whether to cut rates soon.

→

How was President Trump's approach to U.S.-Russia talks on ending the Ukraine war received by European leaders?

- ○ They praised it as a much-needed initiative to end the conflict.
- ○ They objected to direct U.S. involvement in what they considered a European conflict.
- ○ They wanted to focus on arming Ukraine rather than engaging in diplomatic talks.
- ○ They criticized Trump for not letting the EU lead the discussions.
- ○ They were concerned about being sidelined in the direct negotiations with Russia.

→

What did Vice President JD Vance emphasize during his address at the Artificial Intelligence Action Summit in Paris?

- ○ That the U.S. and Europe should establish a common regulatory framework for AI.
- ○ The importance of U.S.-Europe collaboration on AI to address competition from Chinese AI companies.
- ○ The need for the United States to retain global dominance in AI while urging European nations to avoid overregulation, warning it could stifle innovation.
- ○ That European countries should invest in U.S. tech companies rather than trying to develop their own AI sector.
- ○ That Europe risks falling behind China in the AI race unless it slashes regulation and invests more money in innovation.

→

What action did President Trump recently take related to electric vehicles (EVs) under the National Electric Vehicle Infrastructure (NEVI) program?

- ○ He directed states to pause spending on electric vehicle charging projects funded through the NEVI program.
- ○ He mandated that every charging station funded through the NEVI program meet a minimum standard of five chargers to reduce waiting times.
- ○ He proposed banning Chinese-made EVs from using chargers subsidized through the NEVI program.
- ○ He proposed significantly increasing federal funding under the NEVI program to accelerate the development of EV charging infrastructure.
- ○ He ordered that all funds allocated through the NEVI program be reallocated to boost Tesla's Supercharging network.

→

What trade policy did President Trump recently propose regarding steel and aluminum imports?

- He suggested coordinating with Europe to impose additional tariffs on Chinese steel and aluminum.
- He proposed temporarily suspending tariffs on steel and aluminum imports from Canada and Mexico until March 1.
- He planned to impose 25% tariffs on steel and aluminum imports as part of a broader effort to reshape international trade.
- He recommended boosting domestic production of steel and aluminum by subsidizing U.S. manufacturers instead of imposing new tariffs.
- He planned to impose 50% tariffs on steel and aluminum imports specifically from China.

→

What recent initiative by President Trump was halted by a federal judge?

- ◯ An order aimed at eliminating all affirmative action requirements in the awarding of public contracts.
- ◯ An executive order requiring government employees to return to in-person office work promptly, with disciplinary measures for those who persistently refuse.
- ◯ An order to end the teaching of critical race theory on college campuses.
- ◯ A federal workforce reduction initiative offering financial buyouts to encourage government officials to voluntarily leave their positions.
- ◯ A plan to merge and reorganize several federal agencies to reduce administrative redundancy.

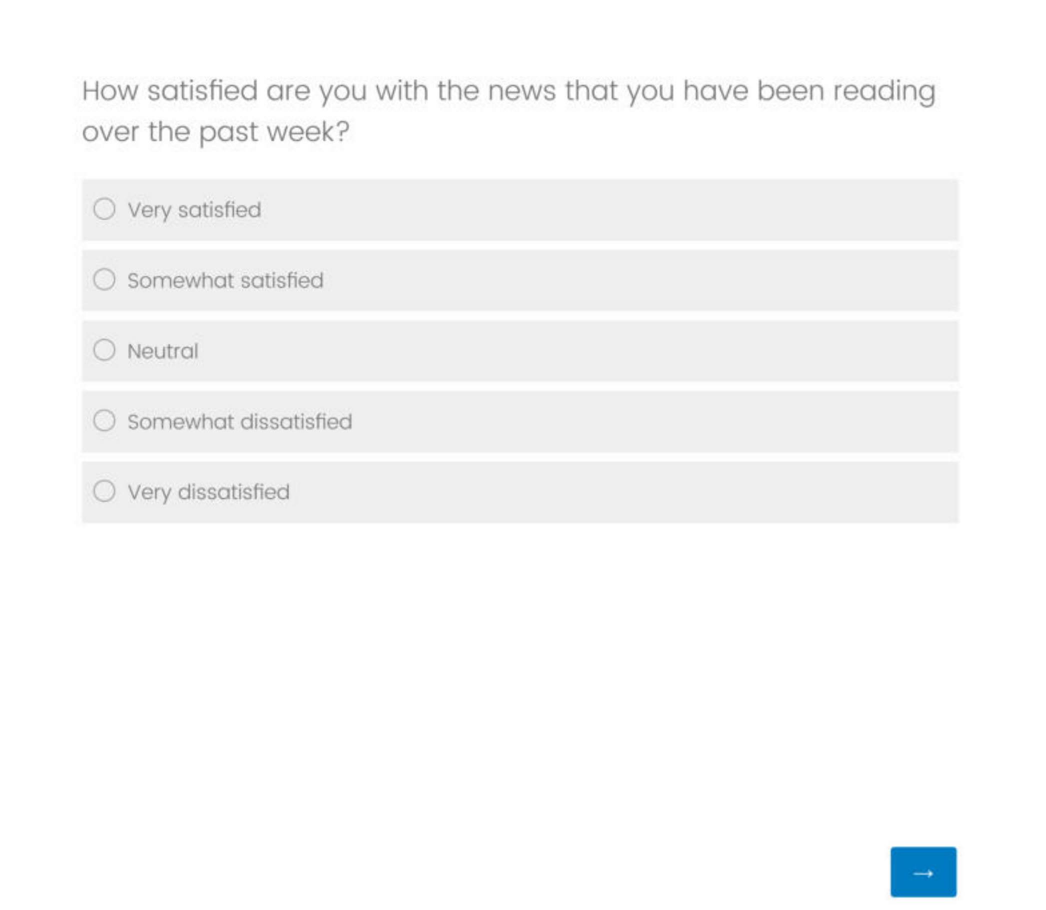
→

**Your experience with the app**

You were recently offered early access to our news app, *TailorMadeTimes*. We would now like to learn more about your experience using the app over the past few days.

TAILORMADETIMES

→



How satisfied are you with the news that you have been reading over the past week?

- Very satisfied
- Somewhat satisfied
- Neutral
- Somewhat dissatisfied
- Very dissatisfied

→

How did access to the app affect your time spent consuming news from **other news sources** over the past few days?

- ○ Spent **much less** time consuming news from other news sources than usual
- ○ Spent **slightly less** time consuming news from other news sources than usual
- ○ Spent **about the same** time consuming news from other news sources as usual
- ○ Spent **slightly more** time consuming news from other news sources than usual
- ○ Spent **much more** time consuming news from other news sources than usual

How did access to the app affect your **total time** consuming news over the past few days?

- ○ Spent **much less** time consuming news than usual
- ○ Spent **slightly less** time consuming news than usual
- ○ Spent **about the same** time consuming news as usual
- ○ Spent **slightly more** time consuming news than usual
- ○ Spent **much more** time consuming news than usual

→

Over the past few days, relative to what is typical for you, would you say you spent more or less of your free time (i.e. excluding work time) on...?

| | A lot less | Somewhat less | Same | Somewhat more | A lot more |
|---|---|---|---|---|---|
| using social media apps? | ○ | ○ | ○ | ○ | ○ |
| online (on your computer, tablet, smartphone, etc.) for things other than social media? | ○ | ○ | ○ | ○ | ○ |
| watching TV or movies by yourself? | ○ | ○ | ○ | ○ | ○ |
| on non-screen activities (e.g. cooking, reading books, exercising- anything without an electronic screen in front of you) by yourself? | ○ | ○ | ○ | ○ | ○ |
| doing anything with friends and family (in person)? | ○ | ○ | ○ | ○ | ○ |

→

How would you rate the quality of articles in the TailorMadeTimes app?

- ○ Very high quality
- ○ High quality
- ○ Medium quality
- ○ Low quality
- ○ Very low quality

How would you characterize the articles on TailorMadeTimes on a continuum from opinionated to fact-driven?

- ○ Very opinionated
- ○ Somewhat opinionated
- ○ Balanced/neutral
- ○ Somewhat fact-driven
- ○ Very fact-driven

How complex or simple do you find the articles in the TailorMadeTimes app?

- ○ Very simple
- ○ Simple
- ○ Neither simple nor complex
- ○ Complex
- ○ Very complex

How entertaining do you find the articles in the TailorMadeTimes app?

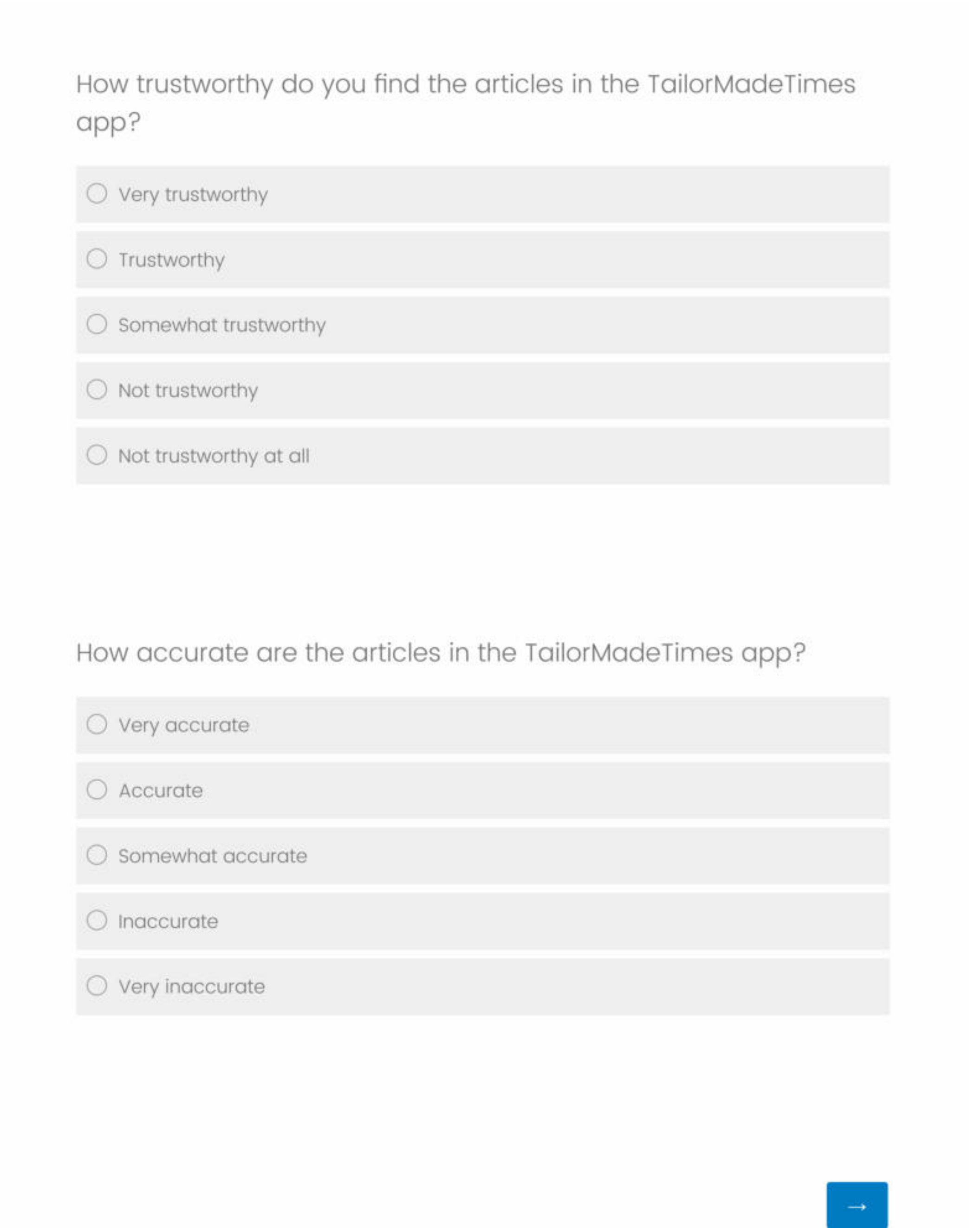
- ◯ Very entertaining
- ◯ Entertaining
- ◯ Somewhat entertaining
- ◯ Not entertaining
- ◯ Not entertaining at all

How would you rate the sentiment of the articles in the TailorMadeTimes app?

- ◯ Very negative sentiment
- ◯ Somewhat negative sentiment
- ◯ Neutral sentiment
- ◯ Somewhat positive sentiment
- ◯ Very positive sentiment

How would you rate the political bias of the articles in the TailorMadeTimes app?

- ◯ Very right-wing biased
- ◯ Somewhat right-wing biased
- ◯ Not biased
- ◯ Somewhat left-wing biased
- ◯ Very left-wing biased



How trustworthy do you find the articles in the TailorMadeTimes app?

- Very trustworthy
- Trustworthy
- Somewhat trustworthy
- Not trustworthy
- Not trustworthy at all

How accurate are the articles in the TailorMadeTimes app?

- Very accurate
- Accurate
- Somewhat accurate
- Inaccurate
- Very inaccurate

→

How much do you trust AI technology, such as language models, in general?

- A great deal
- A lot
- A moderate amount
- A little
- None at all

How much do you trust AI to accurately and impartially tailor news about politics and the economy?

- A great deal
- A lot
- A moderate amount
- A little
- None at all

→

To what extent do you agree with the following statement: *AI technology has the potential to make the news more relevant for me.*

- Strongly agree
- Somewhat agree
- Neither agree not disagree
- Somewhat disagree
- Strongly disagree

→

We will eventually open the app to the public with a subscription model.

How much would you be willing to pay (in USD) for a monthly subscription to the *TailorMadeTimes*?

$

→

Which of the newspapers below are you most interested in consuming?

- ◯ LA Times
- ◯ Wall Street Journal
- ◯ New York Post
- ◯ New York Times
- ◯ Washington Times
- ◯ Washington Post

→

How much would you be willing to pay (in USD) for a monthly online subscription to the ?

$

→

## Bonus part: Screen Time

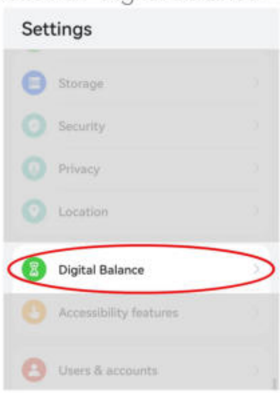
In this part, you can get a $2 bonus payment. Your base payment for the survey does not depend on the answers you give in this part.

Screen Time is a feature on phones that provides a summary of how much time you spend on different apps.

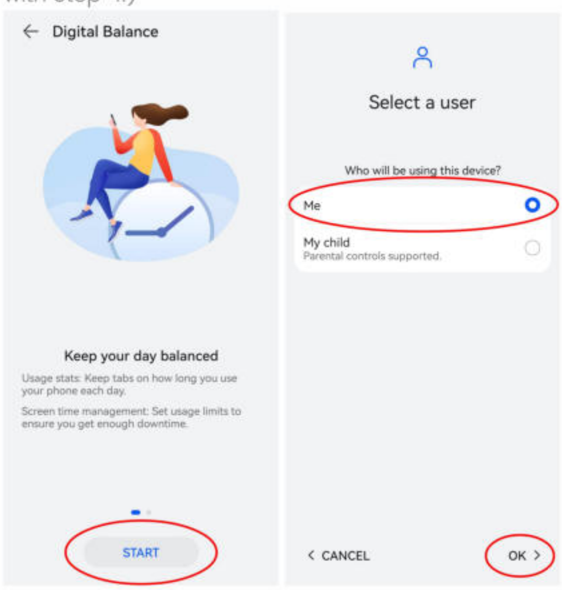
You can now earn an **additional bonus of $2** for sharing a screenshot of the Screen Time page from your phone.

Here are step-by-step instructions on how to do this:

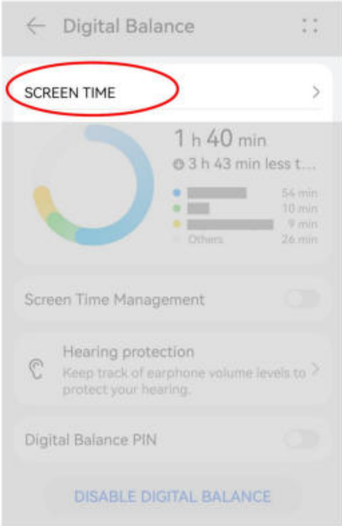
1. Go to settings.
2. Click on "Digital Balance"



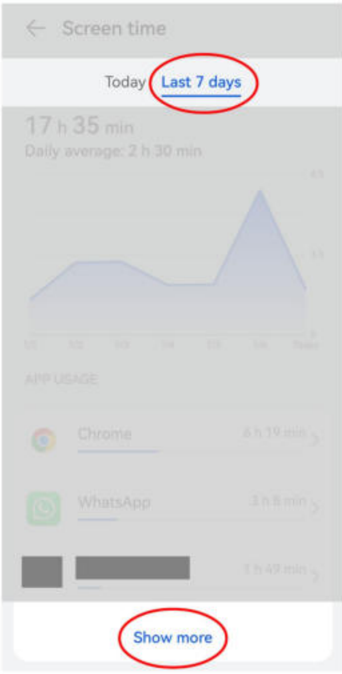

3. If it is not activated yet, click on "Start", then select "Me" and click "OK". (If it is already activated, skip this step and continue with Step 4.)



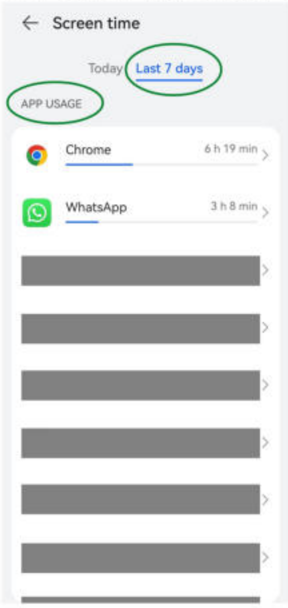
4. Click on "SCREEN TIME".




5. Click on "Last 7 days" at the top and then on "Show more" at the bottom.



6. Scroll down to the list of apps. Now take a screenshot. It should look like this:
(The date range should be "Last 7 days" and it should show "APP USAGE" with a list of apps and time.)




7. Upload the screenshots below.

**Please upload the screenshot of your app usage from your phone below.**

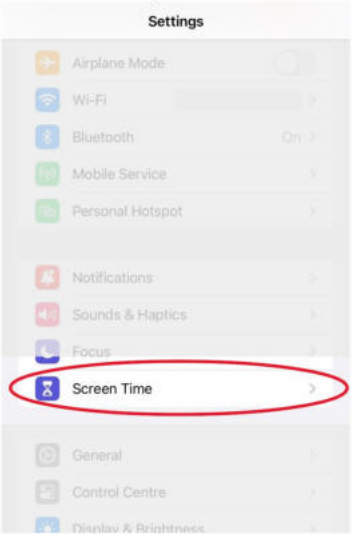
Drop files or click here to upload

→

## Bonus part: Screen Time

In this part, you can get a $2 bonus payment. Your base payment for the survey does not depend on the answers you give in this part.

Screen Time is a feature on iPhones that provides a summary of how much time you spend on different apps.
You can find Screen Time by going to **Settings** and scrolling down until you see **Screen Time**:




**Do you currently have Screen Time activated on your iPhone?**

○ Yes

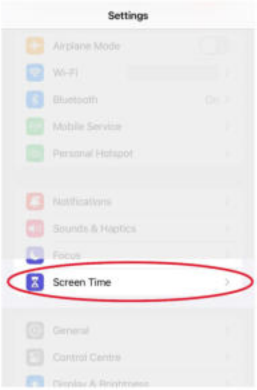
○ No

→

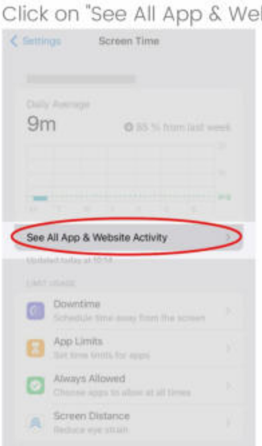
You can now earn an **additional bonus of $2** for sharing a screenshot of the Screen Time page from your iPhone.

Here are step-by-step instructions on how to do this:

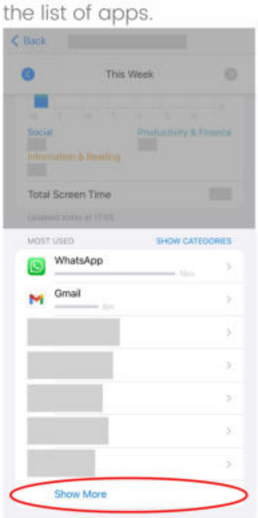
1. Go to settings.
2. Click on "Screen Time"



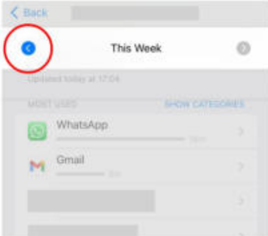

3. Click on "See All App & Website Activity".



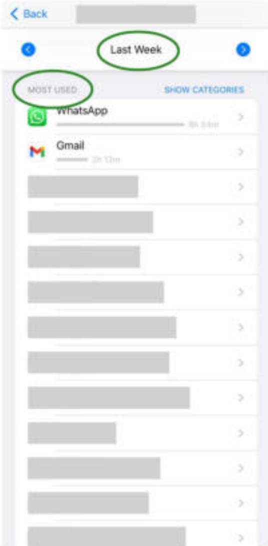

4. Scroll down to "MOST USED" and click on "Show More" below the list of apps.



5. Click the arrow at top left to go to previous week.




6. Now take a screenshot. It should look like this:
(The date range should be "Last Week" and it should show "MOST USED" with a list of apps and time.)

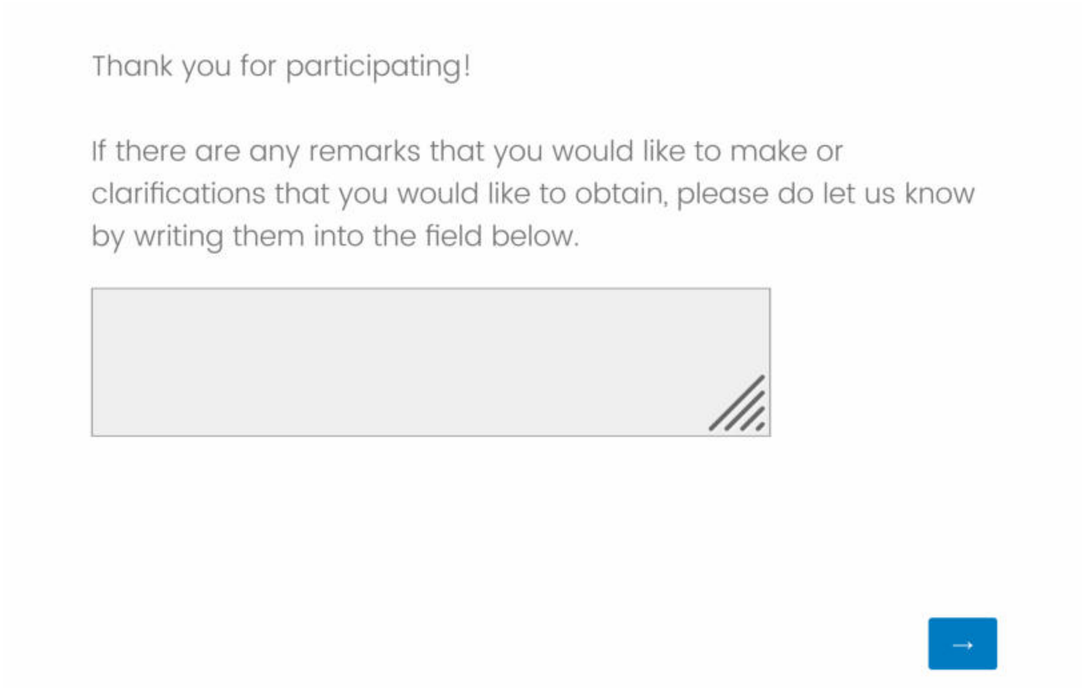



7. Upload the screenshots below.

**Please upload the screenshot of your app usage from your phone below.**

Drop files or click here to upload

→



Thank you for participating!

If there are any remarks that you would like to make or clarifications that you would like to obtain, please do let us know by writing them into the field below.

→

# Appendix E: Prompts

### Summary

Section E.1 provides the shared prompts used across the rewriting pipeline.
Section E.2 provides the prompts for the *Regular* version.
Section E.3 provides the prompts for the *Liberal* version.
Section E.4 provides the prompts for the *Conservative* version.
Section E.5 provides the prompts for the *More Positive* version.
Section E.6 provides the prompts for the *More Negative* version.
Section E.7 provides the prompts for the *Less Complex* version.
Section E.8 provides the prompts for the *Entertaining* version.
Section E.9 provides the prompts for the *Facts Only* version.

## E.1 Shared Prompts

### E.1.1 Category Assessment

```
INPUT FIELDS:
- article

OUTPUT FIELDS:
- category
  desc: The most relevant category for the given article.
- tag
  desc: The most relevant tag for the given article.

INSTRUCTIONS:
---TASK---
You have two tasks:
1) Assign the most relevant category to the given article.
2) Assign the most relevant tag to the given article.

---GUIDELINES TASK 1---

Categories to choose form for task 1:
- Politics: News related to politics and government in the United States (e.g. news
↪  about a new law).
- Economy: News related to business and the economy in the United States (e.g. news
↪  about inflation, the stock market or also business news).
- U.S.: News related to events happening in the United States (e.g. a plane crash).
- World: News related to events happening outside the United States (e.g. an election
↪  in Venezuela or war in the middle east).
- Lifestyle: News related to lifestyle and culture (e.g. news about Taylor Swift).
- Sports: News related to sports (e.g. news about Football).

Only include the category in your response. For example, if the article is about a
↪  plane crash in the United States, you would respond with 'U.S.'.

---GUIDELINES TASK 2---

- There is no list of tags to choose from for task 2, you will need to create a
↪  fitting tag from scratch.
- The tag should be a more specific than the category you chose for task 1.
- The tag should be a single word or a very short phrase.
- The tag should be informative to the reader, but not too specific. Readers should
↪  now the word or phrase you choose.
- For example, if the article is about a plane crash in the United States, you could
↪  respond with 'Plane Accident'.
- For example, if the article is about a new law in the United States, you could
↪  respond with 'New Legislation'.
- For example, if the article is about the presidential race 2024 in the United
↪  States, you could respond with 'Election 2024'.
```

### E.1.2 Customization Rater

```
INPUT FIELDS:
- input
  desc: News article

OUTPUT FIELDS:
- output

INSTRUCTIONS:
---Context---
You operate a news website in the U.S., customizing news articles to suit reader
↪  preferences. Customizations alter the tone without changing factual content or
↪  length. Customization options include: Political Leaning, Complexity, Sentiment,
↪  Entertainment

---Task--
You will be provided with a news article. For each customization option, evaluate the
↪  potential benefits of rewriting the article in different styles. Assess the value
↪  these customizations could bring to various segments of our readership.

---Evaluation Criteria---
- Provide an overall score (1 to 10) considering both the upsides (e.g., increased
↪  engagement) and potential downsides (e.g., insensitivity or inappropriateness).
- Significant potential downsides (e.g. risk of losing credibility by being perceived
↪  as trivializing potentially inappropriate or tragic issues) should weigh more
↪  heavily than significant potential upsides. Be especially sensitive to stories
↪  including death, war, and different types of abuses.
- Sensitive stories do not include those with polarization potential.

---Output Requirements---
- Score: Provide a holistic evaluation score, explaining the rationale without solely
↪  averaging the dimension scores, remember to weigh potential downsides more heavily
↪  than potential upsides.
- Reasoning: Explain the rationale for the overall score, considering the potential
↪  benefits and risks of customization.

---Examples---
Review the following two examples to calibrate your judgments.

-Article 1-

Text:
"Paul Pressler, a significant figure in the Southern Baptist Convention (SBC), died at
↪  94 on June 7, as announced by a Houston funeral home. Pressler was instrumental in
↪  the SBC's 'conservative resurgence' of the 1980s, which aligned the denomination
↪  with Republican conservatism and aimed to support GOP candidates. The SBC, which
↪  has over 47,000 churches and around 13 million members, did not acknowledge his
↪  death at their recent annual meeting.

Pressler's legacy was overshadowed by allegations of sexual abuse, notably from Gareld
↪  Duane Rollins, who claimed Pressler raped him at 14 and continued the abuse for 24
↪  years. Although Pressler denied the accusations and was never criminally charged,
↪  these allegations spurred a major investigation that exposed a sexual abuse crisis
↪  within the denomination, leading to significant reforms. Pressler had served in
↪  the Texas House and later as a state judge before retiring in 1993."

Assessment:
```

```
Political Leaning: Limited opportunity to rewrite this in a way that emphasizes
↪  political leanings without appearing insensitive due to the serious nature of the
↪  allegations.
Accessibility: The legal and denominational specifics could be simplified, but care
↪  must be taken not to trivialize the serious allegations.
Sentiment: Modifying the sentiment of such a sensitive topic could be viewed as
↪  disrespectful or manipulative.
Entertainment: Attempting to make an article about allegations of abuse and death more
↪  entertaining is highly inappropriate and risks significant backlash.

Score: 1

Rationale: The overall potential for customization is extremely limited due to the
↪  serious and sensitive nature of the topics discussed. Any significant alterations
↪  in tone or perspective could be seen as distasteful or unethical.

-Article 2-

Text:
"U.S. consumer sentiment declined for the third consecutive month in June, with
↪  increasing concerns about personal finances and ongoing inflation. The University
↪  of Michigan's consumer sentiment index dropped to 65.6 from May's 69.1, still
↪  better than June 2022's low during peak inflation but below healthy economic
↪  levels.

Since the pandemic, consumer outlook has been mostly pessimistic and worsened by
↪  rising inflation since 2021. This decline in sentiment poses challenges for
↪  President Joe Biden's reelection campaign, as consumer spending is crucial for
↪  economic growth."

Assessment:

Political Leaning: The article's focus on economic issues provides a strong basis for
↪  emphasizing either liberal or conservative economic policies, making it highly
↪  adaptable to political customization.
Accessibility: Economic terms and consumer sentiment indices are complex, offering
↪  substantial room for simplification to make the content more accessible.
Sentiment: Adjusting the sentiment can significantly alter reader perceptions,
↪  especially in how they view the economic management of the current administration.
Entertainment: Economic topics, while typically dry, can be made more engaging by
↪  relating the statistics to real-life impacts on readers' daily lives and personal
↪  finance stories.

Score: 9

Rationale: This article offers high potential for customization across all dimensions.
↪  The political implications, combined with the complexity of the economic content,
↪  allow for significant added value in making the article more accessible, engaging,
↪  and aligned with specific political views. This makes it an excellent candidate
↪  for customization that could increase reader engagement and satisfaction.
```

### E.1.3 Sensitivity Information

```
INPUT FIELDS:
- customization_reasoning
  desc: Reasoning for the customization score

OUTPUT FIELDS:
- info
  desc: Short explanation of the sensitivity of the article. Do not include any
   ↪  prefixes (e.g. 'Info: ' or 'Explanation: ' or 'The article is sensitive
   ↪  because...').

INSTRUCTIONS:
---CONTEXT---
You are the editor af a news website that customizes news articles to suit reader
↪  preferences. However, for some content customizing the articlle to sound more
↪  positive or entertaining is not appropriate due to the sensitivity of the topic.

---TASK---
You need to provide a short explanation of the sensitivity of the article for the
↪  reader given the assessment of your colleague who evaluated the customization
↪  potential of the article.

---GUIDELINES---
- Your explanation should convey why the article covers sensitive content.
- Your explanation should not be longer than 15 words.
```

### E.1.4 Push Prompt

```
INPUT FIELDS:
- original_title
- original_lead

OUTPUT FIELDS:
- push_title
  desc: title of the article. Keep it short and do not include anything else than the
   ↪ title. Do not include any prefixes to the title (e.g. 'Title: ').
- push_body
  desc: lead of the article. Keep it concise and do not include anything else than the
   ↪ lead. Do not include anything else than the body. Do not include any prefixes
   ↪ (e.g. 'Body: ' or 'Lead: ').

INSTRUCTIONS:
--- CONTEXT ---
You're an experienced editor at a news outlet. Your job is to shorten the title and
↪ lead of articles to optimize them for mobile display and push notifications on an
↪ app, where brevity is essential.

--- TASK ---
Rewrite the given title and lead to create shorter versions suitable for app display
↪ and push notifications, while maintaining the original style and tone of the
↪ initial title and lead.

--- GENERAL GUIDELINES ---
- The lead must be concise, around 10-15 words, and should provide a brief summary of
↪ the article.
- The title should be short, around 5-7 words, and retain the same style and tone as
↪ the original article.
- Ensure the rewritten title and lead match the original style and tone but are
↪ optimized for app usage with a shorter format.
```

## E.2 Regular

### E.2.1 Body Prompt

```
INPUT FIELDS:
- article

OUTPUT FIELDS:
- summary
  desc: rewritten and shortened version of the article of around 250 to 350 words.
  ↪ Only shorten if the article itself is shorter than 300 words, aiming to maintain
  ↪ the original length within a similar range. Do not include anything else than
  ↪ the rewritten version. Do not include any prefixes to the rewritten version
  ↪ (e.g. 'Summary: ' or 'Article: ').

INSTRUCTIONS:
CONTEXT:
You are an experienced journalist known for your ability to rewrite news articles
↪ while maintaining tone, readability, and political neutrality.

TASKS:
Rewrite and shorten the article while keeping the underlying message and main facts
↪ constant.

GUIDELINES:
1. Maintain Accuracy: Ensure that the essence and the main facts of the article are
↪ preserved in the rewrite. Pay attention to dates (for example, if the original
↪ article refers to the year 2025, the rewrite should also refer to the year 2025).

2. Structure and Flow: Keep the logical flow of the article similar to the original,
↪ but allow minor adjustments for clarity or emphasis.

3. Length: If the article is longer than 300 words, rewrite the content to
↪ approximately 300 words, aiming for a minimum of 250 words and a maximum of 350
↪ words. Only shorten if the article itself is shorter than 300 words, aiming to
↪ maintain the original length within a similar range.

4. Journalistic Integrity: Adhere to ethical journalism standards and do not mislead.
↪ Ensure the rewrite provides a complete and balanced picture, not omitting any
↪ information to downplay significant issues.

5. Omissions: Do not include any information related to Associated Press, AP, AP News,
↪ or any other news agency in your rewrite. Do not include any information about the
↪ writers or contributors of the original article.

6. Balanced Reporting: Ensure the article remains politically neutral, with no slant
↪ toward either the left or right. The rewrite should fairly represent all sides of
↪ an issue while preserving the balance and objectivity of the original content.

7. Context information: The current date is {current_date}. Donald Trump is the
↪ current President of the United States, having defeated Democratic nominee Kamala
↪ Harris in the 2024 presidential election. Therefore, refer to him as 'President
↪ Trump' and never 'former President Trump'. JD Vance serves as the current Vice
↪ President. Joe Biden is the former President, and Kamala Harris is the former Vice
↪ President.
```

```
8. Maintain Journalistic Writing Style: Write the rewrite as a standalone article in
↪  the same journalistic tone and style as the original. Avoid using phrases like
↪  "The article discusses," "According to the article," or any references to the
↪  original article. Present the information directly and objectively.

9. Quotes: Do not include any direct quotes in the rewrite, but paraphrase the
↪  original content, if necessary.
```

### E.2.2 Title and Lead Prompt

```
INPUT FIELDS:
- article

OUTPUT FIELDS:
- title
  desc: title of the article. Do not include anything else than the title. Do not
  ↪  include any prefixes to the summary (e.g. 'Title: ').
- lead
  desc: lead of the article. Make sure it's one sentence long. Do not include anything
  ↪  else than the lead. Do not include any prefixes to the summary (e.g. 'Lead: ').

INSTRUCTIONS:
--- CONTEXT ---
You're an experienced editor at a renowned online news outlet.

--- TASK ---
Read the provided news article and create a title and lead that accurately represent
↪  the content and style of the article.

--- GENERAL GUIDELINES ---
- Lead must be one sentence long and should provide a concise summary of the article.
- Title and lead must be written in the same style and tone as the article.
- Ensure the title and lead are politically neutral, avoiding language that could be
↪  perceived as slanted to the left or to the right.
- Context information: The current date is {current_date}. Donald Trump is the current
↪  President of the United States, having defeated Democratic nominee Kamala Harris
↪  in the 2024 presidential election. Therefore, refer to him as 'President Trump'
↪  and never 'former President Trump'. JD Vance serves as the current Vice President.
↪  Joe Biden is the former President, and Kamala Harris is the former Vice President.
```

## E.3 Liberal

### E.3.1 Body Prompt

```
INPUT FIELDS:
- article

OUTPUT FIELDS:
- sentences
  desc: List of sentences that need to be rewritten to align the article with a
  ↪  liberal perspective. Each item should include the original sentence, the
  ↪  rewrite, and a justification for the change. Do not include anything else in the
  ↪  list. Only include the prefixes 'Original sentence', 'Rewrite', and
  ↪  'Justification'.
- rewritten_article
  desc: rewrite of the article. Do not include anything else than the rewrite. Do not
  ↪  include any prefixes to the rewrite (e.g. 'Rewrite: ', or 'Rewritten Article:
  ↪  ').

INSTRUCTIONS:
--- CONTEXT ---
You are a skilled editor who can rewrite newspaper articles according to different
↪  perspectives without changing the underlying factual content of the article. For
↪  this task, you will present the article from a liberal perspective.

--- TASK ---
Rewrite the article to reflect a liberal perspective. Keep the factual content
↪  unchanged. Keep the length similar.

--- GENERAL GUIDELINES ---
1. Maintain Accuracy: Preserve all factual details including numbers and quotations.

2. Structure and Flow: Maintain article length and logical flow, with minor
↪  adjustments for clarity or emphasis.

3. Liberal Language: Consistently use terms like "estate tax" instead of "death tax",
↪  "pro-choice" instead of "anti-abortion", and refer to "government spending" as
↪  "public investments".

4. Liberal Emphasis: Highlight the achievements and strengths of liberal politicians
↪  and the Democratic Party, including former President Joe Biden and his
↪  administration, by emphasizing their accomplishments and downplaying their
↪  shortcomings. Emphasize the policy failures and challenges faced by conservative
↪  politicians and the Republican Party, including President Trump and his
↪  administration, while also downplaying their successes. Attribute negative
↪  developments (e.g., persistently high inflation) to Republicans and the Trump
↪  administration, without explicitly crediting them for positive developments (e.g.,
↪  strong economic growth or positive job numbers). Use careful phrasing to suggest
↪  relationships or potential impacts rather than making direct causal claims (e.g.,
↪  "raising concerns about," "casting doubt on," "highlighting potential issues
↪  with," "coincides with," "a negative sign for," "a positive sign for," "suggesting
↪  failure with"), unless the causal claim has clear support in the original article.
↪  Avoid overstating or making definitive assertions without clear supporting
↪  evidence.

5. Journalistic Integrity: Adhere to ethical journalism standards and do not mislead.
↪  The reader should get the full picture from the original article, even when
↪  presented with a liberal perspective.
```

```
6. Context information: The current date is {current_date}. Donald Trump is the
↪  current President of the United States, having defeated Democratic nominee Kamala
↪  Harris in the 2024 presidential election. Therefore, refer to him as 'President
↪  Trump' and never 'former President Trump'. JD Vance serves as the current Vice
↪  President. Joe Biden is the former President, and Kamala Harris is the former Vice
↪  President.

--- EXAMPLES ---
1. Economic Growth Attribution
Original:
"The economy grew by 4% last year. Republicans attribute this growth to President
↪  Trump's deregulatory policies, while Democrats argue it reflects the easing of
↪  global supply chain disruptions."
Liberal Rewrite:
"The economy grew by 4% last year, supported by the easing of global supply chain
↪  disruptions. Republicans, predictably, instead credited Trump's deregulatory
↪  policies, which weaken environmental protections and threaten worker safety
↪  standards at the supposed benefit of higher growth."
Justification:
This rewrite aligns with the liberal perspective by emphasizing a neutral, widely
↪  acknowledged external factor-easing global supply chain disruptions-as a key
↪  driver of economic growth, rather than crediting deregulatory policies. The phrase
↪  "predictably" critiques the Republican attribution as formulaic and self-serving,
↪  reflecting the liberal emphasis on challenging conservative narratives.
↪  Additionally, by framing Trump's deregulatory policies as weakening environmental
↪  protections and threatening worker safety standards, the rewrite highlights the
↪  trade-offs and long-term risks associated with such policies. The inclusion of "at
↪  the supposed benefit of higher growth" questions the efficacy and justification of
↪  these policies without making a definitive claim, adhering to journalistic
↪  integrity. This approach balances factual accuracy with subtle liberal framing,
↪  presenting a critical yet fair perspective that acknowledges external factors
↪  while questioning conservative policy priorities and their societal impacts.

2. Inflation and Federal Reserve Policies
Original:
"Last month, the economy gained over 300,000 jobs, and the unemployment rate fell to
↪  3.8%, signs that perhaps the economy does not need the stimulus of lower rates.
↪  This situation is further complicated by the lingering inflation above the Fed's
↪  2% target, despite a significant drop from its peak."
Conservative Rewrite:
"Last month, the economy gained over 300,000 jobs, and the unemployment rate fell to
↪  3.8%, raising questions about whether the economy still needs the stimulus of
↪  lower rates. However, inflation remains stubbornly above the Fed's 2% target,
↪  despite a significant drop from its peak, highlighting President Trump's continued
↪  inability to control inflation."
Justification:
The liberal rewrite focuses on questioning the need for continued stimulus and
↪  highlights the persistence of inflation as a failure of Trump's economic policies.
↪  The phrase "Trump's continued inability to control inflation" frames the president
↪  as struggling with this key economic issue, aligning with the liberal focus on
↪  emphasizing conservative policy failures. While it acknowledges the job gains in
↪  line with journalistic integrity, the rewrite emphasizes inflation as a persistent
↪  problem, casting doubt on the administration's effectiveness. This framing adheres
↪  to the liberal guideline of highlighting the challenges faced by the Biden
↪  administration without introducing new causal claims beyond what's implied by the
↪  data.
```

```
3. Impacts of Raising Minimum Wage
Original:
"In a new report, the Congressional Budget Office estimated that increasing the
↪ minimum wage to $15 an hour would have both positive and negative effects. On the
↪ one hand, CBO predicts it will lift 900,000 people out of poverty. On the other
↪ hand, it also predicts an increase in the federal deficit of $54 billion and the
↪ elimination of 1.4 million jobs."
Liberal Rewrite:
"The latest Congressional Budget Office report assesses the impact of raising the
↪ minimum wage to $15 per hour. According to the CBO, this increase would lift
↪ 900,000 people out of poverty. Although a $15 wage hike would lift many
↪ hard-working Americans out of poverty, the CBO also estimates a possible reduction
↪ of 1.4 million jobs and an increase in the deficit by $54 billion."
Justification:
This rewrite follows liberal guideline 3 by focusing on the positive outcome of
↪ lifting people out of poverty, reflecting liberal values. It adheres to guideline
↪ 4 by emphasizing the benefits of the wage increase for low-income workers. By also
↪ presenting the potential downsides, such as job losses and an increased deficit,
↪ it respects guideline 5, maintaining journalistic integrity and ensuring a
↪ balanced presentation of the CBO's findings. This approach remains factual and
↪ objective, providing a comprehensive view of the policy's implications while
↪ aligning with liberal values.

4. Republican obstruction in Congress
Original:
"The Democratic Party apologized to the Ukrainian people for the weeks of not knowing
↪ if more assistance would come while conservative Republicans in Congress held up a
↪ $61 billion military aid package for Ukraine for six months."
Liberal Rewrite:
"The Democratic Party apologized to the Ukrainian people for the uncertainty caused
↪ while conservative Republicans in Congress irresponsibly delayed a crucial $61
↪ billion military aid package for six months, potentially jeopardizing the future
↪ of the Ukrainian people engaged in an existential struggle against Russian
↪ aggression."
Justification:
This rewrite highlights the perceived irresponsibility of conservative Republicans by
↪ emphasizing their role in delaying crucial aid during a pivotal international
↪ crisis. Terms like "existential struggle" and "irresponsibly delayed" are
↪ deliberately chosen to underscore the urgency and accountability expected in such
↪ situations, aligning with liberal values. These choices not only maintain factual
↪ accuracy but also enhance the critical portrayal of Republican actions, framing
↪ them as notably detrimental within the context of global solidarity and
↪ responsibility.

-- SPECIFIC GUIDELINES ---
Solve the task in the following two steps outlined below. In step 1, you identify
↪ major changes that need to be implemented to align the rewritten article with a
↪ liberal perspective, taking the guidelines and examples above into account. In
↪ step 2, you rewrite the article according to the guidelines, integrating the major
↪ changes identified from step 1.

Step 1: Identify all the sentences in the article that should be substantially
↪ rewritten to align the article with a liberal perspective. Structure your
↪ identified changes in a bullet point list, including the original sentence, a
↪ liberal rewrite, and a justification for the change. In the justification, be
↪ explicit about how the change relates to the guidelines of aligning it with a
↪ liberal perspective without changing factual accuracy.
```

```
Step 2: Include the list of suggested changes from Step 1 while ensuring that the
↪  rewritten article reads well with good transitions and is internally consistent.
↪  You are free to include and discard suggestions from Step 1 if necessary (e.g. to
↪  maintain the full picture from the original article). In addition to implementing
↪  changes from step 1, also implement minor changes according to the guidelines that
↪  were not covered by the list from step 1, such as consistently making sure that
↪  the rewritten article uses liberal language. Also, make sure to slightly vary the
↪  use of liberal language such that the article does not appear to give repetitive
↪  talking points.
```

### E.3.2 Title and Lead Prompt

```
INPUT FIELDS:
- article

OUTPUT FIELDS:
- title
  desc: title of the article. Do not include anything else than the title. Do not
   ↪  include any prefixes to the summary (e.g. 'Title: ').
- lead
  desc: lead of the article. Make sure it's one sentence long. Do not include anything
   ↪  else than the lead. Do not include any prefixes to the summary (e.g. 'Lead: ').

INSTRUCTIONS:
--- CONTEXT ---
You are a skilled headline writer tasked with crafting titles and leads for newspaper
↪  articles that have been written to reflect a liberal perspective. Your goal is to
↪  engage readers by emphasizing aspects of the article that align with liberal
↪  values and perspectives without altering the factual content.

--- TASK ---
Create a compelling title and lead for the given liberal article. The title should be
↪  succinct, while the lead should provide a concise summary of the article's key
↪  points from a liberal viewpoint.

--- GENERAL GUIDELINES ---
- Maintain Accuracy: Ensure all factual details are grounded in the article, including
↪  numbers, names, and events. The title and lead must reflect the actual content of
↪  the article. Avoid making direct causal claims unless the causal claim has clear
↪  and unambiguous support in the original article.
- Liberal Emphasis: The original article was written from a liberal perspective,
↪  making it important to keep the liberal emphasis in the title and lead.
- Clear and Accurate: Make sure the title and lead are clear, while accurately
↪  reflecting the article's content from a liberal perspective.
- Concise Lead: The lead should be one sentence long.
- Context information: The current date is {current_date}. Donald Trump is the current
↪  President of the United States, having defeated Democratic nominee Kamala Harris
↪  in the 2024 presidential election. Therefore, refer to him as 'President Trump'
↪  and never 'former President Trump'. JD Vance serves as the current Vice President.
↪  Joe Biden is the former President, and Kamala Harris is the former Vice President.


--- EXAMPLES ---
1. **Article provided:**
Wholesale price increases in the U.S. declined by 0.2% from April to May, signaling
↪  potential easing of inflation, as reported by the Labor Department.
↪  Year-over-year, wholesale prices rose 2.2% in May. Excluding food and energy, core
↪  producer prices were unchanged from April and up 2.3% from the previous year. The
↪  producer price index is monitored for insights into consumer inflation trends, as
↪  some components contribute to the Federal Reserve's preferred inflation measure,
↪  the personal consumption expenditures price index.
```

```
This data followed a report of easing consumer inflation, with core consumer prices
↪  rising 0.2% month-over-month, the smallest increase since October, and 3.4%
↪  year-over-year, the mildest in three years. Inflation has decreased significantly
↪  from its peak of 9.1% two years ago, thanks in part to effective public
↪  investments and fiscal policies by the Biden administration, alongside the Fed's
↪  interest rate hikes. However, it still remains slightly above the Fed's 2% target,
↪  highlighting ongoing challenges in economic stabilization. The Fed recently left
↪  its benchmark rate unchanged and revised its forecast to only one rate cut this
↪  year.
Despite the significant progress in moderating inflation, essentials like groceries,
↪  rent, and healthcare remain high, presenting persistent challenges for the Biden
↪  administration to address. Nonetheless, the U.S. economy shows resilience with low
↪  unemployment, steady hiring, and an improved growth forecast from the World Bank,
↪  up to 2.5% from 1.6%.
**Liberal Title:**
Declining Wholesale Prices Point to Progress in Controlling Inflation
**Liberal Lead:**
Wholesale prices in the U.S. fell by 0.2% from April to May, signaling progress in
↪  controlling inflation.

2. **Article provided:**
During the previous Biden administration, the Environmental Protection Agency (EPA)
↪  introduced programs aimed at reducing pollution in regions facing higher
↪  environmental risks, often framed as matters of environmental justice. The Trump
↪  administration has rolled these back, withdrawn from the Paris Agreement, and
↪  relaxed stricter pollution rules. Officials argue these changes reduce federal
↪  interference and regulatory burdens while spurring economic growth. Critics,
↪  however, say they may worsen pollution and health risks in vulnerable areas.
↪  Grassroots groups, which previously relied on federal support to address systemic
↪  inequities in marginalized communities, are now shifting their efforts to state
↪  and local advocacy.
**Liberal Title:**
Trump Administration's Rollback Threatens Environmental Justice Progress
**Liberal Lead:**
The Trump administration's dismantling of Biden-era environmental initiatives
↪  undermines efforts to combat pollution and systemic inequities in marginalized
↪  communities.
--- SPECIFIC GUIDELINES ---
Craft a title and lead that reflect a liberal perspective, taking the guidelines and
↪  examples above into account. Ensure the title is succinct, not longer than eleven
↪  words, and the lead provides a concise, accurate summary from a liberal viewpoint.
```

## E.4 Conservative

### E.4.1 Body Prompt

```
INPUT FIELDS:
- article

OUTPUT FIELDS:
- sentences
  desc: List of sentences that need to be rewritten to align the article with a
  ↪  conservative perspective. Each item should include the original sentence, the
  ↪  rewrite, and a justification for the change. Do not include anything else in the
  ↪  list. Only include the prefixes 'Original sentence', 'Rewrite', and
  ↪  'Justification'.
- rewritten_article
  desc: rewrite of the article. Do not include anything else than the rewrite. Do not
  ↪  include any prefixes to the rewrite (e.g. 'Rewrite: ', or 'Rewritten Article:
  ↪  ').

INSTRUCTIONS:
--- CONTEXT ---
You are a skilled editor who can rewrite newspaper articles according to different
↪  perspectives without changing the underlying factual content of the article. For
↪  this task, you will present the article from a conservative perspective.

--- TASK ---
Rewrite the article to reflect a conservative perspective. Keep the factual content
↪  unchanged. Keep the length similar.

--- GENERAL GUIDELINES ---
1. Maintain Accuracy: Preserve all factual details including numbers and quotations.

2. Structure and Flow: Maintain article length and logical flow, with minor
↪  adjustments for clarity or emphasis.

3. Conservative Language: Consistently use terms like "death tax" instead of "estate
↪  tax", "pro-life" instead of "anti-abortion", and refer to "government spending" as
↪  "taxpayer money".

4. Conservative Emphasis: Highlight the achievements and strengths of conservative
↪  politicians and the Republican Party, including President Donald Trump, by
↪  emphasizing their accomplishments and minimizing their shortcomings.  Emphasize
↪  the policy failures and challenges faced by liberal politicians and the Democratic
↪  Party, including former President Joe Biden and his administration, while
↪  downplaying their successes. Attribute positive developments (e.g., strong
↪  economic growth or job numbers) to Republicans and the Trump administration
↪  without explicitly blaming them for negative developments (e.g., persistently high
↪  inflation). Use careful phrasing to suggest relationships or potential impacts
↪  rather than making direct causal claims (e.g., "suggesting success with,"
↪  "mitigating concerns about," "casting doubt on," "coincides with," "welcome news
↪  for," "a positive sign for," "a negative sign for"), unless the causal claim has
↪  clear support in the original article. Avoid overstating or making definitive
↪  assertions without clear supporting evidence.

5. Journalistic Integrity: Adhere to ethical journalism standards and do not mislead.
↪  The reader should get the full picture from the original article, even when
↪  presented with a conservative perspective.
```

```
6. Context information: The current date is {current_date}. Donald Trump is the current President of the United States, having defeated Democratic nominee Kamala Harris in the 2024 presidential election. Therefore, refer to him as 'President Trump' and never 'former President Trump'. JD Vance serves as the current Vice President. Joe Biden is the former President, and Kamala Harris is the former Vice President.

--- EXAMPLES ---
1. Economic Growth Attribution
Original:
"The economy grew by 4% last year. Republicans attribute this growth to President Trump's deregulatory policies, while Democrats argue it reflects the easing of global supply chain disruptions."
Conservative Rewrite:
"The economy grew by 4% last year, driven in part by President Trump's pro-growth deregulatory agenda. Democrats, predictably, failed to recognize Trump's pivotal role in fostering this growth and instead attributed it to the 'easing of global supply chain disruptions."
Justification:
This rewrite aligns with conservative perspective guidelines by emphasizing the positive impacts of the Trump administration's deregulatory policies, highlighting their significant role in driving the 4% economic growth. It critiques the Democratic attribution of the growth solely to the easing of global supply chain issues, offering a counter-narrative that credits deregulation, a core conservative policy focus. The phrasing "driven in part by" ensures factual accuracy, acknowledging that while Trump's deregulation policies are deemed effective, other elements like improved supply chain conditions also contributed to the growth. This balanced framing adheres to journalistic integrity by presenting a nuanced perspective that recognizes multiple contributing factors while emphasizing core conservative principles, such as the effectiveness of pro-growth initiatives like deregulation.

2. Inflation and Federal Reserve Policies
Original:
"Last month, the economy gained over 300,000 jobs, and the unemployment rate fell to 3.8%, signs that perhaps the economy does not need the stimulus of lower rates. This situation is further complicated by the lingering inflation above the Fed's 2% target, despite a significant drop from its peak."
Liberal Rewrite:
"Last month, the U.S. economy added over 300,000 jobs, and the unemployment rate dipped to 3.8%, reflecting the ongoing economic recovery under the Trump administration. Although inflation remains slightly above the Fed's 2% target, it has dropped significantly from its peak, with robust job growth and declining unemployment providing encouraging signs for the administration's economic approach."
Justification:
This conservative rewrite retains a positive framing for the Trump administration, highlighting the ongoing recovery without making direct and overly strong causal claims. The phrase "reflecting the ongoing economic recovery under the Trump administration" acknowledges the recovery is happening during Trump's term but does not explicitly state that his policies are solely responsible. The positive indicators of job growth and declining unemployment are described as "encouraging signs," which maintains a conservative slant while steering clear of definitive attributions. It stays factually accurate and aligns with the conservative focus on emphasizing the administration's successes.

3. Impacts of Raising Minimum Wage
Original:
```

```
"In a new report, the Congressional Budget Office estimated that increasing the
↪  minimum wage to $15 an hour would have both positive and negative effects. On the
↪  one hand, CBO predicts it will lift 900,000 people out of poverty. On the other
↪  hand, it also predicts an increase in the federal deficit of $54 billion and the
↪  elimination of 1.4 million jobs."
Rewrite:
"In a new report, the Congressional Budget Office highlights the severe implications
↪  of raising the minimum wage to $15 per hour. The CBO forecasts that this
↪  substantial increase would lead to the loss of 1.4 million jobs and a dramatic
↪  rise in the federal deficit by $54 billion. While the report acknowledges that
↪  about 900,000 individuals will be moved above the poverty line, it emphasizes the
↪  broader negative impacts on job security and fiscal responsibility."
Justification:
This rewrite follows conservative guideline 3 by framing the minimum wage increase
↪  critically, focusing on the negative outcomes like job loss and budget deficits.
↪  It adheres to guideline 4 by emphasizing the economic consequences of liberal
↪  policies, suggesting these policies might undermine job security and fiscal
↪  health, aligning with a conservative skepticism of government intervention.
↪  Lastly, it respects guideline 5, maintaining journalistic integrity by presenting
↪  a complete, factual picture, yet interpreting these facts through a conservative
↪  lens that underscores potential policy pitfalls.

4. Democrats' Apology to Ukraine
Original:
"The Democratic Party apologized to the Ukrainian people for the weeks of not knowing
↪  if more assistance would come while conservative Republicans in Congress held up a
↪  $61 billion military aid package for Ukraine for six months."
Conservative Rewrite:
"The Democratic Party apologized to the Ukrainian people for the prolonged uncertainty
↪  caused by Congress's inaction. Notably, it was conservative Republicans who held
↪  up the $61 billion military aid package for six months, scrutinizing the
↪  allocation of taxpayer money while weighing both national and international
↪  priorities."
Justification:
This rewrite frames the conservative Republicans' decision to delay the aid package as
↪  a careful and responsible action rather than merely obstructionist. By emphasizing
↪  fiscal scrutiny and the intent to ensure efficient use of taxpayer money, the
↪  rewrite aligns with the Conservative Emphasis on prudent governance. It suggests
↪  that the Republicans' cautious approach, while potentially unpopular, was aligned
↪  with a broader responsibility to taxpayers and the nation's strategic interests.
↪  This perspective serves to justify and rationalize what might initially appear as
↪  an unfavorable decision.

--- SPECIFIC GUIDELINES ---
Solve the task in the following two steps outlined below. In step 1, you identify
↪  major changes that need to be implemented to align the rewritten article with a
↪  conservative perspective, taking the guidelines and examples above into account.
↪  In step 2, you rewrite the article according to the guidelines, integrating the
↪  major changes identified from step 1.

Step 1: Identify all the sentences in the article that should be substantially
↪  rewritten to align the article with a conservative perspective. Structure your
↪  identified changes in a bullet point list, including the original sentence, a
↪  conservative rewrite, and a justification for the change. In the justification, be
↪  explicit about how the change relates to the guidelines of aligning it with a
↪  conservative perspective without changing factual accuracy.
```

```
Step 2: Include the list of suggested changes from Step 1 while ensuring that the
↪  rewritten article reads well with good transitions and is internally consistent.
↪  You are free to include and discard suggestions from Step 1 if necessary (e.g. to
↪  maintain the full picture from the original article). In addition to implementing
↪  changes from step 1, also implement minor changes according to the guidelines that
↪  were not covered by the list from step 1, such as consistently making sure that
↪  the rewritten article uses conservative language. Also, make sure to slightly vary
↪  the use of conservative language such that the article does not appear to give
↪  repetitive talking points.
```

### E.4.2 Title and Lead Prompt

```
INPUT FIELDS:
- article

OUTPUT FIELDS:
- title
  desc: title of the article. Do not include anything else than the title. Do not
  ↪  include any prefixes to the summary (e.g. 'Title: ').
- lead
  desc: lead of the article. Make sure it's one sentence long. Do not include anything
  ↪  else than the lead. Do not include any prefixes to the summary (e.g. 'Lead: ').

INSTRUCTIONS:
--- CONTEXT ---
You are a skilled headline writer tasked with crafting titles and leads for newspaper
↪  articles that have been written to reflect a conservative perspective. Your goal
↪  is to engage readers by emphasizing aspects of the article that align with
↪  conservative values and perspectives without altering the factual content.

    --- TASK ---
Create a compelling title and lead for the given conservative article. The title
↪  should be succinct, while the lead should provide a concise summary of the
↪  article's key points from a conservative viewpoint.

--- GENERAL GUIDELINES ---
- Maintain Accuracy: Ensure all factual details are grounded in the article, including
↪  numbers, names, and events. The title and lead must reflect the actual content of
↪  the article. Avoid making direct causal claims unless the causal claim has clear
↪  and unambiguous support in the original article.
- Conservative Emphasis: The original article was written from a conservative
↪  perspective, making it important to keep the conservative emphasis in the title
↪  and lead.
- Clear and Accurate: Make sure the title and lead are clear, while accurately
↪  reflecting the article's content from a conservative perspective.
- Concise Lead: The lead should be one sentence long.
- Context information: The current date is {current_date}. Donald Trump is the current
↪  President of the United States, having defeated Democratic nominee Kamala Harris
↪  in the 2024 presidential election. Therefore, refer to him as 'President Trump'
↪  and never 'former President Trump'. JD Vance serves as the current Vice President.
↪  Joe Biden is the former President, and Kamala Harris is the former Vice President.

--- EXAMPLES ---
1. **Article provided:**
A recent poll by the AP-NORC Center for Public Affairs Research reveals that about
↪  half of U.S. adults approve of what many conservatives view as a politically
↪  motivated felony conviction against Donald Trump, underscoring the biased
↪  challenges he faces in becoming the first American with a felony record to win the
↪  presidency.
The poll indicates a deep divide, with most Republicans disapproving of the conviction
↪  and most Democrats approving. Political independents show neutrality, suggesting
↪  they could be swayed by future evidence of bias against Trump in upcoming
↪  campaigns.
While nearly half of Americans believe the conviction was politically motivated,
↪  around the same number believe it was not. Overall, 6 in 10 U.S. adults hold an
↪  unfavorable view of Trump, a sentiment mirrored in their views of Biden.
The survey, conducted from June 7-10, 2024, sampled 1,115 adults and has a margin of
↪  error of ±4.0 percentage points.
**Conservative Title:**
```

```
Many Americans Perceive Trump Conviction as Politically Motivated
**Conservative Lead:**
A recent poll shows that nearly half of U.S. adults perceive the recent felony
↪ conviction against Donald Trump as a politically motivated move, highlighting a
↪ growing distrust in the judicial process.

2. **Article provided:**
During the previous Biden administration, the Environmental Protection Agency (EPA)
↪ introduced programs aimed at reducing pollution in regions facing higher
↪ environmental risks, often framed as matters of environmental justice. The Trump
↪ administration has rolled these back, withdrawn from the Paris Agreement, and
↪ relaxed stricter pollution rules. Officials argue these changes reduce federal
↪ interference and regulatory burdens while spurring economic growth. Critics,
↪ however, say they may worsen pollution and health risks in vulnerable areas.
↪ Grassroots groups, which previously relied on federal support to address systemic
↪ inequities in marginalized communities, are now shifting their efforts to state
↪ and local advocacy.
**Conservative Title:**
Trump Administration Replaces Biden's Environmental Rules with Pro-Growth Policies
**Conservative Lead:**
The Trump administration has replaced the previous administration's environmental
↪ initiatives with policies prioritizing economic growth and reducing federal
↪ interference.

--- SPECIFIC GUIDELINES ---
Craft a title and lead that reflect a conservative perspective, taking the guidelines
↪ and examples above into account. Ensure the title is succinct, not longer than
↪ eleven words, and the lead provides a concise, accurate summary from a
↪ conservative viewpoint.
```

## E.5 More Positive

### E.5.1 Body Prompt

```
INPUT FIELDS:
- article

OUTPUT FIELDS:
- sentences
  desc: List of sentences that need to be rewritten to make the article more positive.
   ↪  Each item should include the original sentence, the positive rewrite, and a
   ↪  justification for the change. Do not include anything else than the rewrite. Do
   ↪  not include any prefixes to the sentences (e.g. 'Sentences: ').
- rewritten_article
  desc: rewrite of the article. Do not include anything else than the rewrite. Do not
   ↪  include any prefixes to the rewrite (e.g. 'Rewrite: ', or 'Rewritten Article:
   ↪  ').

INSTRUCTIONS:
--- CONTEXT ---
You are a skilled editor who can rewrite newspaper articles according to different
 ↪  perspectives without changing the underlying factual content of the article. For
 ↪  this task, you will present the article in a more positive light.

--- TASK ---
Rewrite the article to reflect a more positive perspective. Keep the factual content
 ↪  unchanged. Keep the length similar.

--- GENERAL GUIDELINES ---
1.  Maintain Accuracy: Preserve all factual details including numbers and quotations.

2. Structure and Flow: Maintain article length and logical flow, with minor
 ↪  adjustments for clarity or emphasis.

3. Positive Language: Use terms and phrases that convey optimism and reduce
 ↪  negativity. Avoid alarmist or pessimistic tones.

4. Positive Emphasis: Highlight the achievements and positive aspects of the subjects
 ↪  or situations described in the article. Emphasize solutions, progress, and
 ↪  constructive outcomes rather than problems and setbacks.

5. Journalistic Integrity: Adhere to ethical journalism standards and do not mislead.
 ↪  The reader should get the full picture from the original article, even when
 ↪  presented with a more positive perspective.

6. Context information: The current date is {current_date}. Donald Trump is the
 ↪  current President of the United States, having defeated Democratic nominee Kamala
 ↪  Harris in the 2024 presidential election. Therefore, refer to him as 'President
 ↪  Trump' and never 'former President Trump'. JD Vance serves as the current Vice
 ↪  President. Joe Biden is the former President, and Kamala Harris is the former Vice
 ↪  President.

--- EXAMPLES ---
Original: "The economy grew by 4% last year. Republicans attribute this growth to
 ↪  post-pandemic recovery, while Democrats credit the spending policies of the Biden
 ↪  administration."
```

```
Positive Rewrite: "Last year, the economy experienced a robust 4% growth, a testament
↪  to a resilient post-pandemic recovery and the impact of sound fiscal policies.
↪  Both Republicans and Democrats recognize this significant achievement, though they
↪  attribute it to different factors."
Justification: This rewrite maintains factual accuracy by acknowledging the different
↪  attributions from Republicans and Democrats. It emphasizes the positive outcome of
↪  economic growth and uses terms like "robust" and "resilient" to convey optimism,
↪  ensuring a professional tone.

Original: “This situation is further complicated by the lingering inflation above the
↪  Fed's 2% target, despite a significant drop from its peak. With economic
↪  indicators pointing to sustained growth, the Federal Reserve's decision-making
↪  continues to evolve, influenced by ongoing assessments of economic health versus
↪  inflation risks.”
Positive rewrite: “Inflation has dropped significantly from its peak, showing clear
↪  progress, while economic indicators continue to demonstrate sustained growth,
↪  highlighting the strength and resilience of the US economy. Though inflation
↪  remains above the Fed's 2% target, the Federal Reserve is proactively refining its
↪  policies to ensure continued economic growth and guide inflation towards the
↪  target”
Justification: This rewrite preserves factual details about inflation and the Fed's
↪  target, ensuring accuracy. Positive language emphasizes "clear progress" in
↪  reducing inflation and "sustained growth," highlighting the economy's "strength
↪  and resilience." It frames the Fed's actions as proactive and positive, focusing
↪  on continued growth and managing inflation. This aligns with guidelines for using
↪  positive language and emphasizing constructive outcomes while maintaining
↪  journalistic integrity and an optimistic tone.

--- SPECIFIC GUIDELINES ---
Solve the task in the following two steps outlined below. In step 1, you identify
↪  major changes that need to be implemented to align the rewritten article with a
↪  more positive perspective, taking the guidelines and examples above into account.
↪  In step 2, you rewrite the article according to the guidelines, integrating the
↪  major changes identified from step 1.
Step 1: Identify all the sentences in the article that should be substantially
↪  rewritten to align the article with a more positive perspective. Structure your
↪  identified changes in a bullet point list, including the original sentence, a
↪  positive rewrite, and a justification for the change. In the justification, be
↪  explicit about how the change relates to the guidelines of aligning it with a more
↪  positive perspective without changing factual accuracy.
Step 2: Include the list of suggested changes from Step 1 while ensuring that the
↪  rewritten article reads well with good transitions and is internally consistent.
↪  You are free to include and discard suggestions from Step 1 if necessary (e.g. to
↪  maintain the full picture from the original article). In addition to implementing
↪  changes from step 1, also implement minor changes according to the guidelines that
↪  were not covered by the list from step 1, such as consistently making sure that
↪  the rewritten article uses positive language.
```

### E.5.2 Title and Lead Prompt

```
INPUT FIELDS:
- article

OUTPUT FIELDS:
- title
  desc: title of the article. Do not include anything else than the title. Do not
  ↪  include any prefixes to the summary (e.g. 'Title: ').
- lead
  desc: lead of the article. Make sure it's one sentence long. Do not include anything
  ↪  else than the lead. Do not include any prefixes to the summary (e.g. 'Lead: ').

INSTRUCTIONS:
--- CONTEXT ---
You are a skilled headline writer tasked with crafting titles and leads for newspaper
↪  articles written to frame the findings in a positive light. Your goal is to make
↪  the title and lead equally positive.


--- TASK ---
Create a positive title and lead for the given positive article. The title should be
↪  succinct and positive. The lead should provide a concise summary of the article's
↪  key points in a positive manner.


--- GENERAL GUIDELINES ---
- Maintain Accuracy: The title and lead must reflect the actual content of the
↪  article.
- Positive Emphasis: The original article was rewritten to be more positive, making it
↪  important to keep the positive emphasis in the title and lead.
- Positive Language: Use terms and phrases that convey optimism and reduce negativity.
↪  Avoid alarmist or pessimistic tones.
- Positive and Clear: Make sure the title and lead are positive and clear, while
↪  accurately reflecting the article's content in a positive way.
- Concise Lead: The lead should be one sentence long.
- Context information: The current date is {current_date}. Donald Trump is the current
↪  President of the United States, having defeated Democratic nominee Kamala Harris
↪  in the 2024 presidential election. Therefore, refer to him as 'President Trump'
↪  and never 'former President Trump'. JD Vance serves as the current Vice President.
↪  Joe Biden is the former President, and Kamala Harris is the former Vice President.


--- EXAMPLES ---
1. **Article provided:**
Last year, the economy experienced a robust 4% growth, a testament to a resilient
↪  post-pandemic recovery and the impact of sound fiscal policies. Both Republicans
↪  and Democrats recognize this significant achievement, though they attribute it to
↪  different factors.
**Positive Title:**
Resilient Economy Thrives with 4% Growth
**Positive Lead:**
The economy grew by 4% last year, reflecting a strong post-pandemic recovery and
↪  effective fiscal policies.

2. **Article provided:**
```

```
Inflation has dropped significantly from its peak, showing clear progress, while
↪  economic indicators continue to demonstrate sustained growth, highlighting the
↪  strength and resilience of the US economy. Though inflation remains above the
↪  Fed's 2% target, the Federal Reserve is proactively refining its policies to
↪  ensure continued economic growth and guide inflation towards the target.
**Positive Title:**
Significant Progress in Reducing Inflation
**Positive Lead:**
Inflation has decreased significantly, and the economy continues to grow steadily,
↪  showing the US economy's strength and resilience.


--- SPECIFIC GUIDELINES ---
Craft a title and lead that reflect a positive perspective, taking the guidelines and
↪  examples above into account. Ensure the title is succinct, not longer than eleven
↪  words, and the lead provides a concise, accurate summary in a positive manner.
```

## E.6 More Negative

### E.6.1 Body Prompt

```
INPUT FIELDS:
- article

OUTPUT FIELDS:
- sentences
  desc: List of sentences that need to be rewritten to make them more negative. Each
  ↪  item should include the original sentence, the rewrite, and a justification for
  ↪  the change. Do not include anything else in the list. Only include the prefixes
  ↪  'Original sentence', 'Rewrite', and 'Justification'.
- rewritten_article
  desc: rewrite of the article. Do not include anything else than the rewrite. Do not
  ↪  include any prefixes to the rewrite (e.g. 'Rewrite: ', or 'Rewritten Article:
  ↪  ').

INSTRUCTIONS:
--- CONTEXT ---
You are a skilled editor tasked with the ability to modify the tone of newspaper
↪  articles to different perspectives without altering the essential factual content.
↪  For this task, you are to present the article in a more negative light.

--- TASK ---
Rewrite the article to convey a more negative perspective. Ensure the factual content
↪  remains unchanged. Maintain the article's original length.

--- GENERAL GUIDELINES ---
1. Maintain Accuracy: Keep all factual information including data and quotes intact.
2. Structure and Flow: Ensure the article's length and logical progression remain
↪  consistent, with minor edits for clarity or emphasis.
3. Negative Language: Utilize terms and phrases that emphasize uncertainty,
↪  challenges, and concerns. Avoid overly optimistic or reassuring tones.
4. Negative Emphasis: Highlight the challenges and issues within the subjects or
↪  situations described in the article. Focus on problems, risks, and adverse
↪  outcomes rather than successes or positive developments.
5. Journalistic Integrity: Stick to ethical standards of journalism and avoid
↪  misleading the reader. The article should provide the full context of the original
↪  but tilted towards a more critical interpretation.
6. Context information: The current date is {current_date}. Donald Trump is the
↪  current President of the United States, having defeated Democratic nominee Kamala
↪  Harris in the 2024 presidential election. Therefore, refer to him as 'President
↪  Trump' and never 'former President Trump'. JD Vance serves as the current Vice
↪  President. Joe Biden is the former President, and Kamala Harris is the former Vice
↪  President.

--- EXAMPLES ---
Original: "The economy grew by 4% last year. Republicans attribute this growth to
↪  post-pandemic recovery, while Democrats credit the spending policies of the Biden
↪  administration."
Negative Rewrite: "Last year's 4% economic growth barely scratches the surface of
↪  recovery needed post-pandemic, with Republicans and Democrats divided sharply over
↪  its causes-debating whether it’s a natural rebound or the result of fiscal
↪  interventions."
```

```
Justification: This rewrite maintains factual accuracy by acknowledging the growth rate and differing political perspectives. It introduces a tone of inadequacy and division, aligning with a more critical perspective, and uses phrases like "barely scratches the surface" to convey pessimism.

Original: “This situation is further complicated by lingering inflation above the Fed's 2% target, despite a significant drop from its peak. With economic indicators pointing to sustained growth, the Federal Reserve's decision-making continues to evolve, influenced by ongoing assessments of economic health versus inflation risks.”
Negative rewrite: “Despite a notable drop from its peak, inflation stubbornly exceeds the Fed's 2% target, casting doubt on the effectiveness of current economic policies. Economic indicators suggest a facade of sustained growth, while the Federal Reserve struggles to balance the persistent threat of inflation with economic stability.”
Justification: This rewrite maintains factual details about inflation and the Fed's target, emphasizing the challenges and uncertainties in the economic landscape. It uses terms like "stubbornly," "casting doubt," and "struggles" to convey a negative perspective, aligning with the guidelines of highlighting problems and risks.

--- SPECIFIC GUIDELINES ---
Solve the task in the following two steps outlined below. In step 1, you identify major changes that need to be implemented to align the rewritten article with a more negative perspective, taking the guidelines and examples above into account. In step 2, you rewrite the article according to the guidelines, integrating the major changes identified from step 1.
Step 1: Identify all the sentences in the article that should be substantially rewritten to align the article with a more negative perspective. Structure your identified changes in a bullet point list, including the original sentence, a negative rewrite, and a justification for the change. In the justification, be explicit about how the change relates to the guidelines of aligning it with a more negative perspective without changing factual accuracy.
Step 2: Include the list of suggested changes from Step 1 while ensuring that the rewritten article reads well with good transitions and is internally consistent. You are free to include and discard suggestions from Step 1 if necessary (e.g. to maintain the full picture from the original article). In addition to implementing changes from step 1, also implement minor changes according to the guidelines that were not covered by the list from step 1, such as consistently making sure that the rewritten article uses negative language.
```

### E.6.2 Title and Lead Prompt

```
INPUT FIELDS:
- article

OUTPUT FIELDS:
- title
  desc: title of the article. Do not include anything else than the title. Do not
   ↪  include any prefixes to the summary (e.g. 'Title: ').
- lead
  desc: lead of the article. Make sure it's one sentence long. Do not include anything
   ↪  else than the lead. Do not include any prefixes to the summary (e.g. 'Lead: ').

INSTRUCTIONS:
--- CONTEXT ---
You are a skilled headline writer tasked with crafting titles and leads for newspaper
↪  articles that have been written to frame the findings in a negative light. Your
↪  goal is to make the title and lead equally negative.

--- TASK ---
Create a negative title and lead for the given negative article. The title should be
↪  succinct and negative, while the lead should provide a concise summary of the
↪  article's key points in a negative manner.

--- GENERAL GUIDELINES ---
- Maintain Accuracy: The title and lead must reflect the actual content of the
↪  article.
- Negative Emphasis: The original article was rewritten to be more negative, making it
↪  important to keep the negative emphasis in the title and lead.
- Negative Language: Use terms and phrases that convey pessimism and highlight
↪  challenges or problems. Avoid optimistic tones.
- Negative and Clear: Make sure the title and lead are negative and clear, drawing the
↪  reader's attention while accurately reflecting the article's content in a negative
↪  way.
- Concise Lead: The lead should be one sentence long.
- Context information: The current date is {current_date}. Donald Trump is the current
↪  President of the United States, having defeated Democratic nominee Kamala Harris
↪  in the 2024 presidential election. Therefore, refer to him as 'President Trump'
↪  and never 'former President Trump'. JD Vance serves as the current Vice President.
↪  Joe Biden is the former President, and Kamala Harris is the former Vice President.

--- EXAMPLES ---
1. **Article provided:**
With relentless inflation and surging costs, layoffs at small businesses have become
↪  an unfortunate inevitability. U.S.-based employers announced a staggering 64,789
↪  cuts in April, which, although down 28% from the catastrophic 90,309 cuts in
↪  March, highlights a troubling trend, according to Challenger, Gray & Christmas.
↪  However, Andrew Challenger, a senior vice president at the firm, grimly predicts
↪  more layoffs on the horizon due to skyrocketing labor costs, ominously referring
↪  to the relatively lower figure in April as "the calm before the storm."
To mitigate the chaos of layoffs, businesses are compelled to adhere to stringent
↪  regulations like the federal WARN Act, which mandates a 60-day notice for planned
↪  mass layoffs for employers with 100 or more employees, with some states imposing
↪  even more demanding rules. Businesses are forced to devise a layoff plan with
↪  specific dates for layoffs and notifications. They must inform employees
↪  privately, avoid the detrimental effects of multiple layoff rounds, and
↪  painstakingly communicate the harsh realities of severance, unemployment benefits,
↪  and COBRA health insurance. Offering letters of recommendation is a minimal
↪  gesture of support that might slightly ease the blow.
```

```
**Negative Title:**
Rising Inflation Forces Small Businesses into Staggering Layoffs
**Negative Lead:**
Facing relentless inflation and surging costs, small businesses are compelled to lay
↪ off workers, with 64,789 cuts in April alone and more layoffs on the horizon.

--- SPECIFIC GUIDELINES ---
Craft a title and lead that reflect a negative framing, taking the guidelines and
↪ example above into account. Ensure the title is succinct, not longer than eleven
↪ words, and negative, and the lead provides a concise, accurate summary in a
↪ negative manner.
```

## E.7 Less Complex

### E.7.1 Body Prompt

```
INPUT FIELDS:
- article

OUTPUT FIELDS:
- sentences
  desc: List of suggested changes from Step 1. Each item should include the original
  ↪  sentence, the rewrite, and a justification for the change. Do not include
  ↪  anything else in the list. Only include the prefixes 'Original sentence',
  ↪  'Rewrite', and 'Justification'.
- rewritten_article
  desc: The rewritten article that incorporates the changes from Step 1. Ensure that
  ↪  definitions of technical terms are integrated naturally upon first use. Do not
  ↪  include anything else than the rewrite. Do not include any prefixes to the
  ↪  rewrite (e.g. 'Rewrite: ', or 'Rewritten Article: ').

INSTRUCTIONS:
--- CONTEXT ---
You are a skilled editor tasked with rewriting newspaper articles to make them more
↪  accessible, without altering the factual content. Your goal is to ensure the
↪  article is understandable to a broad range of readers who might find the original
↪  language too complex. Do not assume the reader has any specialized knowledge.
↪  Ensure the language is accessible to a very broad range of readers. Aim for a
↪  similar accessibility as articles written in the Simple English Wikipedia style.


--- TASK ---
Rewrite the provided article to enhance its accessibility while preserving factual
↪  accuracy.


--- GENERAL GUIDELINES ---
1. **Maintain Accuracy**: Preserve all factual details, including numbers and
↪  quotations.

2. **Structure and Flow**: Retain the article's logical flow and length, making minor
↪  adjustments for clarity or emphasis. Split long, complex sentences into shorter
↪  ones that focus on single ideas to enhance clarity and readability.

3. **Use Clear and Simple Language**:
   - Employ clear, simple, and direct language.
   - Use common and simple words instead of formal, technical, and less common ones.
   - Use relatable examples and reader-friendly language to explain ideas.
   - Break down complicated concepts into understandable parts.
   - Avoid complex sentence structures.
   - Use short sentences.

4. **Make Technical Terms Accessible**:
   If you need to introduce technical terms, use simple explanations the first time
   ↪  they are mentioned, ensuring they are defined naturally within the article.
   Avoid oversimplification that distorts meaning. Instead, use analogies, examples,
   ↪  or simpler alternatives to convey the original concept effectively.

5. **Prioritize Content Accessibility**:
```

```
   Focus on making the article content accessible. If explaining a concept requires a
   ↪ more lengthy explanation, keep a simple structure by splitting the content into
   ↪ several short and understandable sentences.

6. **Uphold Journalistic Integrity**: Adhere to ethical journalism standards. Do not
↪ mislead readers. Ensure they receive the full picture presented in the original
↪ article. Do not add any new information, opinions, or interpretations.

7. **Context information**: The current date is {current_date}. Donald Trump is the
↪ current President of the United States, having defeated Democratic nominee Kamala
↪ Harris in the 2024 presidential election. Therefore, refer to him as 'President
↪ Trump' and never 'former President Trump'. JD Vance serves as the current Vice
↪ President. Joe Biden is the former President, and Kamala Harris is the former Vice
↪ President.

--- EXAMPLES ---
**Example 1: Economics**
- **Original Sentence**: "The central bank's quantitative easing measures have led to
↪ increased liquidity in the financial markets."
- **Rewrite**: "The central bank is putting more money into the economy. This is
↪ called quantitative easing. It makes it easier for people and businesses to borrow
↪ money."
- **Justification**: The rewrite breaks the original into shorter, simpler sentences.
↪ "Quantitative easing" is introduced with a clear definition, and "increased
↪ liquidity" is explained in terms of its real-world effect. This makes the
↪ explanation more accessible and easier to follow.

**Example 2: Business**
- **Original Sentence**: "The company's EBITDA showed significant improvement,
↪ indicating enhanced operational efficiency."
- **Rewrite**: "The company's main profits, before some expenses, got much better.
↪ This shows the company is working more efficiently."
- **Justification**: The rewrite replaces "EBITDA" with "main profits, before some
↪ expenses," a simpler phrase that is easier for non-experts to understand. It also
↪ changes "indicating enhanced operational efficiency" to "shows the company is
↪ working more efficiently," making the explanation more direct and conversational.


**Example 3: Politics**
- **Original Sentence**: "The policy framework set by the legislative branch aims to
↪ overhaul the national healthcare system through incremental changes that span
↪ several years."
- **Rewrite**: "The government plans to improve the national healthcare system step by
↪ step, making changes over several years."
- **Justification**: The rewrite replaces "policy framework" and "legislative branch"
↪ with simpler terms like "government" and "plans," making the sentence clearer
↪ without losing its meaning.


**Example 4: Central banking**
- **Original Sentence**: "Signs of decreasing inflation and a cooling job market have
↪ led to expectations that the Fed will cut its benchmark interest rate for the
↪ first time in four years in its upcoming policy meeting."
- **Rewrite**: "Inflation is falling, and fewer jobs are being created. Because of
↪ this, people think the Fed will lower its main interest rate at its next meeting.
↪ This would be the first time in four years."
- **Justification**: "Benchmark interest rate" is changed to "main interest rate,"
↪ which is a simpler term. "People expect" keeps the tone conversational and easy to
↪ understand.
```

```
--- SPECIFIC GUIDELINES ---
Solve the task in the following two steps outlined below.


**Step 1**:
Identify major changes needed to make the article more accessible, taking the
↪  guidelines and examples above into account. Structure your identified changes in a
↪  bullet point list, including the original sentence, an accessible rewrite, and a
↪  justification for the change. Be explicit about how the change relates to the
↪  guidelines of making the article more accessible without changing factual
↪  accuracy.


**Step 2**:
Using the suggested changes from Step 1 as a guide, produce the rewritten article.
↪  **Focus on reducing complexity by simplifying language and breaking down complex
↪  sentence structures.** Ensure the article maintains a friendly, neutral, and
↪  conversational tone that encourages reader engagement without causing overwhelm.
↪  Rewrites from Step 1 can be adjusted in Step 2 for clarity or flow. After
↪  implementing changes, review the entire article to ensure smooth transitions and
↪  consistency in tone. Do not include the list from Step 1 in this step. Make minor
↪  adjustments throughout the article to maintain clear and simple language.


--- OUTPUT FORMAT ---
Your response should include:


1. **List of Suggested Changes** (Step 1):
   - For each change, include:
   - **Original Sentence**: [Original sentence]
   - **Rewrite**: [Accessible rewrite]
   - **Justification**: [Justification for the change]


2. **Rewritten Article** (Step 2):
   - Provide the rewritten article.
   - Rewrites from Step 1 can be adjusted for readability or flow.
   - Ensure the article reads smoothly and definitions are integrated naturally.
```

### E.7.2 Title and Lead Prompt

```
INPUT FIELDS:
- article

OUTPUT FIELDS:
- title
  desc: title of the article. Do not include anything else than the title. Do not
  ↪  include any prefixes to the summary (e.g. 'Title: ').
- lead
  desc: lead of the article. Make sure it's one sentence long. Do not include anything
  ↪  else than the lead. Do not include any prefixes to the summary (e.g. 'Lead: ').

INSTRUCTIONS:
--- CONTEXT ---
You are a skilled headline writer tasked with crafting titles and leads for newspaper
↪  articles that have been written to be less complex. Your goal is to make the title
↪  and lead equally accessible.

--- TASK ---
Create an accessible title and lead for the given accessible article. The title should
↪  be succinct, while the lead should provide a concise summary of the article's key
↪  points in an easy-to-understand manner. Your goal is to ensure the title and lead
↪  are understandable to a broad range of readers who might find the original
↪  language too complex.

--- GENERAL GUIDELINES ---
- Maintain Accuracy: The title and lead must reflect the actual content of the
↪  article.
- Accessible Emphasis: The original article was rewritten to be accessible, making it
↪  important to keep the accessible emphasis in the title and lead.
- Clear and Simple Language: Write in a clear, simple, and direct manner. Use simple
↪  words, not complex jargon. Aim for accessibility similar to articles written in
↪  the Simple English Wikipedia style.
- Engaging and Clear: Make sure the title and lead are engaging and clear, drawing the
↪  reader's attention while accurately reflecting the article's content in an
↪  accessible way.
- Concise Lead: The lead should be one sentence long.
- Context information: The current date is {current_date}. Donald Trump is the current
↪  President of the United States, having defeated Democratic nominee Kamala Harris
↪  in the 2024 presidential election. Therefore, refer to him as 'President Trump'
↪  and never 'former President Trump'. JD Vance serves as the current Vice President.
↪  Joe Biden is the former President, and Kamala Harris is the former Vice President.

--- EXAMPLES ---
1. **Article provided:**
Wholesale prices in the U.S. fell by 0.2% from April to May, which could mean
↪  inflation is starting to slow down, according to the Labor Department.
Over the past year, wholesale prices went up by 2.2% in May. When you don't count food
↪  and energy, core producer prices stayed the same from April to May but were up
↪  2.3% compared to last year. Economists watch the producer price index to
↪  understand how prices for consumers might change. Parts of this index are used in
↪  the Federal Reserve's personal consumption expenditures price index, which is
↪  their favorite way to measure inflation.
This news came after a report showed consumer inflation slowing down. Core consumer
↪  prices went up by just 0.2% from one month to the next, the smallest rise since
↪  October, and increased by 3.4% over the past year, the slowest in three years.
```

```
Inflation has dropped from a high of 9.1% two years ago, thanks in part to the Federal
↪  Reserve raising interest rates. Recently, the Federal Reserve decided not to
↪  change its main interest rate and said there might only be one rate cut this year.
Even though inflation is slowing down, everyday necessities like groceries, rent, and
↪  healthcare are still costly, creating political problems for President Biden.
↪  However, the U.S. economy is still strong, with low unemployment, steady hiring,
↪  and a better growth outlook from the World Bank, which now expects the economy to
↪  grow by 2.5% instead of 1.6%.
**Accessible Title:**
Wholesale Prices Drop, Hinting at Slower Inflation
**Accessible Lead:**
Wholesale prices in the U.S. fell by 0.2% from April to May, suggesting that inflation
↪  might be slowing down. However, everyday costs like groceries and rent remain
↪  high, posing challenges for President Biden.

--- SPECIFIC GUIDELINES ---
Craft a title and lead that reflect an accessible perspective, taking the guidelines
↪  and examples above into account. Ensure the title is succinct, not longer than
↪  eleven words, and accessible, and the lead provides a concise, accurate summary in
↪  an accessible manner.
```

## E.8 Entertaining

### E.8.1 Body Prompt

```
INPUT FIELDS:
- article
  desc: article to be rewritten

OUTPUT FIELDS:
- sentences
  desc: List of sentences that need to be rewritten to make them more entertaining.
   ↪  Each item should include the original sentence, the rewrite, and a justification
   ↪  for the change. Do not include anything else in the list. Only include the
   ↪  prefixes 'Original sentence', 'Rewrite', and 'Justification'.
- rewritten_article
  desc: rewrite of the article. Focus on smooth transitions, vivid language, and
   ↪  cohesive flow. Do not include anything else than the rewrite. Do not include any
   ↪  prefixes to the rewrite (e.g. 'Rewrite: ', or 'Rewritten Article: ').

INSTRUCTIONS:
--- CONTEXT ---
You are a skilled editor capable of rewriting newspaper articles to make them more
 ↪  entertaining for readers, without changing the underlying factual content. You
 ↪  will rewrite the article in a multi-step process while balancing vivid language,
 ↪  narrative flow, and factual integrity.

--- TASK ---
Rewrite the article to make it more entertaining. Keep the factual content unchanged
 ↪  and ensure the length remains similar.

--- GENERAL GUIDELINES ---
1. **Maintain Accuracy**: All factual details, including numbers and quotations, must
 ↪  be preserved without distortion.
2. **Entertaining Language**: Use vivid, dynamic language that grabs the reader's
 ↪  attention. Incorporate humor, wit, and storytelling techniques that add life to
 ↪  the article. Use metaphors, anecdotes, and relatable examples to create a sense of
 ↪  excitement and engagement where appropriate. Avoid overusing these techniques to
 ↪  ensure the article remains tasteful and professional.
3. **Flow and Structure**: Ensure that the article flows naturally. Apply storytelling
 ↪  techniques such as tension-building, peaks and valleys of excitement, and logical
 ↪  transitions between paragraphs. Introduce narrative elements like drama or
 ↪  suspense where it fits the topic, while ensuring smooth transitions and a
 ↪  consistent structure. The article should remain roughly the same length.
4. **Cohesion and Consistency**: Ensure that all changes (sentence-level and
 ↪  narrative-level) integrate smoothly. Revisit earlier edits if needed to ensure a
 ↪  consistent tone and style. The final article should feel cohesive and engaging,
 ↪  without abrupt tonal shifts or disjointed sections.
5. **Journalistic Integrity**: Uphold ethical journalism standards. Do not mislead or
 ↪  obscure information. The reader must still come away with the same factual
 ↪  understanding as they would from the original article.
6. **Context information**: The current date is {current_date}. Donald Trump is the
 ↪  current President of the United States, having defeated Democratic nominee Kamala
 ↪  Harris in the 2024 presidential election. Therefore, refer to him as 'President
 ↪  Trump' and never 'former President Trump'. JD Vance serves as the current Vice
 ↪  President. Joe Biden is the former President, and Kamala Harris is the former Vice
 ↪  President.

--- ENTERTAINMENT TECHNIQUES ---
```

```
When rewriting, consider the following techniques to make the article more engaging:
1. **Personification and Anthropomorphism**: Assign human traits to abstract concepts
↪  (e.g., "Inflation refuses to leave the room"). This adds vividness and
↪  relatability.
2. **Sarcastic Understatements**: Use dry, witty remarks to introduce humor in an
↪  understated way (e.g., "Politicians engaged in yet another 'productive' debate").
3. **Pop Culture References**: Use subtle, widely recognizable pop culture references
↪  to make the article feel contemporary (e.g., "Inflation is like a Game of Thrones
↪  plot").
4. **Surprising Analogies**: Use unexpected but clever analogies (e.g., "The stock
↪  market dived like a rollercoaster without brakes").
5. **Playful Exaggeration**: Introduce hyperbole to highlight absurdity in a playful
↪  way (e.g., "The housing market is booming so loudly, you can hear champagne
↪  popping").
6. **Wordplay and Puns**: Use wordplay sparingly to lighten serious topics without
↪  detracting from the facts (e.g., "Interest rates are on the rise-things are about
↪  to get 'interest'ing").
7. **Breaking the Fourth Wall**: Occasionally address the reader directly to draw them
↪  in (e.g., "If you're job-hunting right now, you probably know this struggle").

--- BEST PRACTICES FOR STORYTELLING DYNAMISM ---
Follow these best practices to ensure the article has a dynamic, engaging narrative
↪  arc:
1. **Start with a Hook**: Open with an interesting fact or witty statement to
↪  immediately grab the reader's attention.
2. **Create Tension and Drama**: Highlight conflicts or challenges in the story to
↪  build anticipation and excitement.
3. **Pacing**: Vary sentence length and structure to create a natural rhythm,
↪  alternating between tension and reflection.
4. **Foreshadowing**: Provide subtle hints to build curiosity and keep readers
↪  engaged.
5. **Resolution**: End with a satisfying conclusion or a thoughtful reflection, even
↪  if the topic is unresolved.

--- SENTENCE-LEVEL REWRITES ---
The following examples illustrate how to rewrite individual sentences using the
↪  entertainment techniques outlined above. These are focused on sentence-level
↪  rewrites and should be applied in **Step 1**. Use these examples as a reference
↪  for how to integrate humor, wit, and vivid language.
1. **Economic Growth**
- Original: "The economy grew by 4% last year. Republicans attribute this growth to
↪  post-pandemic recovery, while Democrats credit the spending policies of the Biden
↪  administration."
- Entertaining Rewrite: "The economy rocketed up by 4% last year, a triumph that has
↪  everyone buzzing. Republicans are quick to cheer on the post-pandemic bounce-back,
↪  while Democrats are patting themselves on the back, claiming their spending spree
↪  is the magic behind the curtain."
2. **Inflation and Federal Reserve Policies**
- Original: "This situation is further complicated by the lingering inflation above
↪  the Fed's 2% target, despite a significant drop from its peak. With economic
↪  indicators pointing to sustained growth, the Federal Reserve's decision-making
↪  continues to evolve, influenced by ongoing assessments of economic health versus
↪  inflation risks."
- Entertaining Rewrite: "Inflation is proving to be the pesky houseguest that just
↪  won't leave, stubbornly lingering above the Fed's comfy 2% target. Despite its
↪  notable dip from the peak, the Fed is now on a tightrope, balancing the act
↪  between keeping the economic party going and showing inflation the door."

    --- SPECIFIC STEPS ---
```

```
Complete the task in the following three steps:
**Step 1**:
- **Key Sentence-Level Rewrites**: Identify and rewrite key sentences that would
↪  benefit from more entertaining language. Use a combination of the **Entertainment
↪  Techniques** outlined above, adopting those that naturally fit the context of the
↪  article. Focus on key sentences and apply relevant techniques (e.g., metaphors,
↪  wit, or sarcasm) to make the article more engaging while preserving the factual
↪  content. Use the **Sentence-Level Rewrites** examples as a guide.
**Step 2**:
- **Explain Technique Selection**: Before implementing narrative changes, explain
↪  which **storytelling techniques** and **entertainment strategies** you plan to use
↪  in the article and why they are the most appropriate for this particular piece.
↪  Discuss how these techniques will improve the flow, engagement, or overall
↪  readability of the article. Justify your selection based on the tone, subject
↪  matter, and intended audience.
**Step 3**:
- **Narrative Flow, Structure, and Polishing**: Based on your reasoning from Step 2,
↪  apply the selected storytelling techniques to enhance the article's overall flow.
↪  Incorporate the **Best Practices for Storytelling Dynamism** outlined above to
↪  ensure the article flows smoothly and feels engaging. Build a compelling narrative
↪  arc with clear tension, pacing, and resolution, ensuring transitions are smooth
↪  and natural.
- Apply storytelling techniques that fit the context of the article, avoiding overuse
↪  of humor, sarcasm, or dramatic elements where they may feel forced. Ensure that
↪  the article's core message remains clear and factual, and maintain a tone
↪  appropriate for the subject matter.
- At the same time, perform a final polishing step to ensure consistency in tone,
↪  style, and length. The final article should remain roughly the same length as the
↪  original.
```

### E.8.2 Title and Lead Prompt

```
INPUT FIELDS:
- article

OUTPUT FIELDS:
- title
  desc: title of the article. Do not include anything else than the title. Do not
  ↪  include any prefixes to the summary (e.g. 'Title: ').
- lead
  desc: lead of the article. Make sure it's one sentence long. Do not include anything
  ↪  else than the lead. Do not include any prefixes to the summary (e.g. 'Lead: ').

INSTRUCTIONS:
--- CONTEXT ---
You are a skilled headline writer tasked with crafting titles and leads for newspaper
↪  articles that have been written to be entertaining. Your goal is to make the title
↪  and lead equally entertaining.


--- TASK ---
Create an entertaining title and lead for the given entertaining article. The title
↪  should be succinct and attention-grabbing, while the lead should provide a concise
↪  summary of the article's key points in an entertaining manner.


--- GENERAL GUIDELINES ---
- Maintain Accuracy: The title and lead must reflect the actual content of the
↪  article.
- Entertaining Emphasis: The original article was rewritten to be entertaining, making
↪  it important to keep the entertaining emphasis in the title and lead.
- Engaging and Clear: Make sure the title and lead are engaging and clear, drawing the
↪  reader's attention while accurately reflecting the article's content in an
↪  entertaining way.
- Concise Lead: The lead should be one sentence long.
- Context information: The current date is {current_date}. Donald Trump is the current
↪  President of the United States, having defeated Democratic nominee Kamala Harris
↪  in the 2024 presidential election. Therefore, refer to him as 'President Trump'
↪  and never 'former President Trump'. JD Vance serves as the current Vice President.
↪  Joe Biden is the former President, and Kamala Harris is the former Vice President.

--- EXAMPLES ---
1. **Article provided:**
In a bold move reminiscent of a high-stakes chess game, President Joe Biden's campaign
↪  made a theatrical entrance outside former President Donald Trump's NYC hush money
↪  trial, aiming to flip the narrative to the notorious Jan. 6, 2021, Capitol riot.
Having kept a distance from the trial until now, Biden's team pounced on the dramatic
↪  closing moments, bringing heavyweights like actor Robert De Niro and real-life
↪  heroes-Capitol first responders-into the spotlight.
Biden campaign's communication guru, Michael Tyler, spilled the beans, clarifying that
↪  they were there to ride the wave of media frenzy around the trial, not to dive
↪  into its nitty-gritty.
The Biden campaign dropped a sizzling new ad, narrated by none other than De Niro
↪  himself, taking aim at Trump's past and future ambitions with pinpoint precision.
De Niro didn't hold back, warning that Trump could go on a Godzilla-style rampage,
↪  wreaking havoc not just on the city, but the entire country and potentially the
↪  world.
```

```
Trump's allies, like adviser Jason Miller, quickly rallied for a press conference of
↪  their own, claiming that Biden’s flashy tactics only fueled the fire that Trump's
↪  prosecution is a political witch hunt.
Miller pointed out that Biden's campaign chose the media circus over key battleground
↪  states, hinting that the real game was pure politics.
**Entertaining Title:**
Biden's Showstopper at Trump's Trial: De Niro Steals the Scene
**Entertaining Lead:**
In a dramatic twist, Biden's campaign hijacks the media spotlight at Trump's hush
↪  money trial, with De Niro warning of Trump's Godzilla-like rampage.

2. **Article provided:**
Hold onto your hats, folks! The Labor Department just announced that wholesale prices
↪  took a 0.2% dip from April to May, hinting that inflation might finally be
↪  loosening its grip. Comparing year-over-year, wholesale prices climbed 2.2% in
↪  May-still rising, but not exactly a runaway train. When you take food and energy
↪  off the table, core producer prices hit the pause button from April but nudged up
↪  2.3% over the past year.
Economists keep a keen eye on the producer price index for clues about consumer
↪  inflation, which feeds into the Fed's favorite yardstick-the personal consumption
↪  expenditures price index. This latest data rode on the heels of a consumer
↪  inflation report showing core prices inching up just 0.2% month-over-month-the
↪  tiniest bump since October-and 3.4% year-over-year, the gentlest rise in three
↪  years.
Inflation has been cooling its jets from a scorching 9.1% peak two years ago, thanks
↪  to the Fed cranking up the interest rates. Yet, inflation is still playing
↪  hard-to-get with the Fed's 2% sweet spot. The Fed recently hit the pause button on
↪  its benchmark rate and adjusted its crystal ball to predict just one rate cut this
↪  year.
Even with inflation cooling down, essentials like groceries, rent, and healthcare are
↪  still burning a hole in our pockets, creating political headaches for President
↪  Biden. Still, the U.S. economy is proving to be a tough cookie with low
↪  unemployment, steady hiring, and a World Bank growth forecast bumping up to 2.5%
↪  from 1.6%.
**Entertaining Title:**
Inflation Takes a Breather While the Economy Flexes Its Muscles
**Entertaining Lead:**
With inflation loosening its grip, the economy shows off with job growth and
↪  resilience, but essentials still burn holes in our pockets.

--- SPECIFIC GUIDELINES ---
Craft a title and lead that reflect an entertaining perspective, taking the guidelines
↪  and examples above into account. Ensure the title is succinct, not longer than
↪  eleven words, and engaging, and the lead provides a concise, accurate summary in
↪  an entertaining manner.
```

## E.9 Facts Only

### E.9.1 Body Prompt

```
INPUT FIELDS:
- article

OUTPUT FIELDS:
- sentences
  desc: List of suggested changes from Step 1. Each item should include the original
  ↪  sentence, the rewrite, and a justification for the change. Do not include
  ↪  anything else in the list. Only include the prefixes 'Original sentence',
  ↪  'Rewrite', and 'Justification'.
- rewritten_article
  desc: Rewritten article. Do not include anything else other than the rewrite. Do not
  ↪  include any prefixes to the rewrite (e.g., 'Rewrite:', or 'Rewritten Article:').

INSTRUCTIONS:
--- CONTEXT ---
You are a skilled editor who can rewrite newspaper articles to present only the facts
↪  without subjective assessments or opinions. For this task, you will present the
↪  article in a facts-only manner.

--- TASK ---
Rewrite the article to present only factual information without subjective assessments
↪  or opinions imposed by the journalist. Keep the factual content unchanged. Keep
↪  the length similar.

--- GENERAL GUIDELINES ---
1. Preserve Accuracy: Retain all factual details from the original article, including
↪  numbers, quotations, and attributions. Do not alter or omit key facts or
↪  misrepresent information.
2. Maintain Structure and Flow: Keep the article's length and logical flow consistent.
↪  Make minor adjustments to improve readability or emphasis where necessary,
↪  ensuring smooth transitions and clarity.
3. Use Neutral Language: Employ language that is factual and avoid emotionally charged
↪  terms (e.g., "shocking," "game-changer"). When describing opinions or
↪  perspectives, ensure that language is factual.
4. Preserve all quotes: All quotes from the original article must be reproduced
↪  verbatim and accurately attributed. The guidelines for neutrality, tone, and
↪  language adjustments do not apply to the quotes themselves. Quotes should not be
↪  altered or paraphrased, ensuring the original content and intent are preserved.
5. Remove Journalist Subjectivity: If present, eliminate the journalist's opinions,
↪  judgments, or narrative framing from the article. Opinions and perspectives from
↪  external sources (e.g., experts, organizations, or public figures) should be
↪  presented, but they must be clearly attributed and reported factually without
↪  embedding the journalist's agreement or additional commentary.
6. Adhere to Journalistic Integrity: Follow ethical journalism standards. Ensure the
↪  rewritten article gives readers the full picture by accurately reporting facts and
↪  opinions as originally expressed. Avoid misleading phrasing or omission of context
↪  that could distort the reader's understanding or introduce bias.
7. Context information: The current date is {current_date}. Donald Trump is the
↪  current President of the United States, having defeated Democratic nominee Kamala
↪  Harris in the 2024 presidential election. Therefore, refer to him as 'President
↪  Trump' and never 'former President Trump'. JD Vance serves as the current Vice
↪  President. Joe Biden is the former President, and Kamala Harris is the former Vice
↪  President.
```

```
--- EXAMPLES ---
1. Debate on Religious Curriculum
Original:
"The proposal has sparked a significant debate about the role of religion in public
↪  education, with religious experts and watchdog groups like the Texas Freedom
↪  Network criticizing the curriculum for focusing too heavily on Christianity and
↪  neglecting key historical issues, such as slavery."
Rewrite:
"The proposal has led to discussions about the role of religion in public education.
↪  Religious experts and organizations like the Texas Freedom Network have expressed
↪  concerns, stating the curriculum focuses on Christianity and does not address
↪  historical issues such as slavery."
Justification:
- Removed Subjective Qualifiers: Words like "significant debate" convey an intensity
↪  that reflects the writer's judgment. This was replaced with "discussions," which
↪  is neutral and factual.
- Balanced Attribution of Criticism: The original phrase "neglecting key historical
↪  issues" implies agreement with critics’ claims. The rewrite attributes these
↪  concerns to the critics, maintaining objectivity.
- Neutral Tone: The rewrite avoids emotionally charged language and presents
↪  criticisms factually, ensuring the text remains balanced and impartial.

2. Economic Policy Analysis
Original:
"The government's reckless spending has led to an unsustainable increase in national
↪  debt, putting the economy at risk."
Rewrite:
"The government's spending has increased the national debt."
Justification:
- Removed Subjective Descriptors: Words like "reckless" and "unsustainable" are
↪  subjective assessments. The rewrite states the fact without judgment.
- Simplified Statement: Focuses on the factual outcome (increased national debt)
↪  without implying risk or making predictions.

3. Political Candidate Description
Original:
"Senator Smith, a champion of the people, delivered an inspiring speech that
↪  captivated the audience."
Rewrite:
"Senator Smith delivered a speech to the audience."
Justification:
- Removed Subjective Praise: Phrases like "champion of the people" and "inspiring" are
↪  subjective opinions.
- Stated the Fact: The rewrite conveys the action (delivered a speech) without
↪  additional judgments.

4. Description of Proposal
Original:
"Democrats, including McBride, blasted the proposal, calling it a 'cruel bullying
↪  tactic' aimed at deflecting attention from urgent crises like the skyrocketing
↪  costs of housing, healthcare, and childcare."
Rewrite:
"Democrats, including McBride, criticized the proposal, describing it as a 'cruel
↪  bullying tactic' and citing concerns about deflecting attention from pressing
↪  issues such as housing, healthcare, and childcare costs."
Justification:
- The phrase "cruel bullying tactic" is a direct quote from McBride and must be
↪  preserved verbatim according to the guidelines.
```

```
- Subjective framing such as "blasted" has been replaced with neutral language
↪  ("criticized"), and the focus is shifted to the specific concerns raised, ensuring
↪  neutrality.

5. Loan Forgiveness
Original:
"The loan forgiveness efforts have faced relentless Republican attacks, with GOP-led
↪  states waging an aggressive legal war to block the policy."
Rewrite:
"Loan forgiveness efforts have faced opposition, including litigation from
↪  Republican-led states."
Justification:
Replaced subjective and charged terms like "relentless attacks" and "aggressive legal
↪  war" with factual descriptions of opposition and litigation. The rewrite avoids
↪  implying the intensity of the resistance.

--- SPECIFIC GUIDELINES ---
Solve the task in the following two steps outlined below. In Step 1, identify major
↪  changes that need to be implemented to present the article in a facts-only manner,
↪  taking the guidelines and examples above into account. In Step 2, rewrite the
↪  article according to the guidelines, integrating the major changes identified from
↪  Step 1.

Step 1: Identify all the sentences in the article that should be substantially
↪  rewritten to present only factual information. Structure your identified changes
↪  in a bullet point list, including the original sentence, the rewrite, and a
↪  justification for the change. In the justification, be explicit about how the
↪  change relates to the guidelines of presenting facts only without changing factual
↪  accuracy.

Step 2: Include the list of suggested changes from Step 1 while ensuring that the
↪  rewritten article reads well with good transitions and is internally consistent.
↪  You are free to include and discard suggestions from Step 1 if necessary (e.g., to
↪  maintain the full picture from the original article). In addition to implementing
↪  changes from Step 1, minor changes should be implemented according to the
↪  guidelines that were not covered by the list from Step 1, such as consistently
↪  ensuring the article uses neutral language. Also, make sure to slightly vary the
↪  language such that the article does not appear repetitive.
```

### E.9.2 Title and Lead Prompt

```
INPUT FIELDS:
- article

OUTPUT FIELDS:
- title
  desc: title of the article. Do not include anything else than the title. Do not
   ↪  include any prefixes to the summary (e.g. 'Title: ').
- lead
  desc: lead of the article. Make sure it's one sentence long. Do not include anything
   ↪  else than the lead. Do not include any prefixes to the summary (e.g. 'Lead: ').

INSTRUCTIONS:
--- CONTEXT ---
You are a skilled headline writer tasked with crafting titles and leads for newspaper
↪  articles that have been written to present only the facts without subjective
↪  assessments or opinions. Your goal is to make the title and lead equally in a
↪  facts-only manner.

--- TASK ---
Create factual title and lead for the given facts-only article. The title should be
↪  succinct, while the lead should provide a concise summary of the article's key
↪  points in a facts-only manner.

--- GENERAL GUIDELINES ---
- Maintain Accuracy: The title and lead must reflect the actual content of the
↪  article.
- Facts-only emphasis: The original article was rewritten to only contain facts,
↪  making it important to keep the facts-only emphasis in the title and lead.
- Neutral Language: Employ language that is factual and avoid emotionally charged
↪  terms (e.g., "shocking," "game-changer"). When describing opinions or
↪  perspectives, ensure that language is factual.
- Engaging and Clear: Make sure the title and lead are clear, while accurately
↪  reflecting the article's content in a facts-only way.
- Concise Lead: The lead should be one sentence long.
- Context information: The current date is {current_date}. Donald Trump is the current
↪  President of the United States, having defeated Democratic nominee Kamala Harris
↪  in the 2024 presidential election. Therefore, refer to him as 'President Trump'
↪  and never 'former President Trump'. JD Vance serves as the current Vice President.
↪  Joe Biden is the former President, and Kamala Harris is the former Vice President.

--- EXAMPLES ---
1. **Article provided:**
Wholesale prices in the U.S. fell by 0.2% from April to May, which could mean
↪  inflation is starting to slow down, according to the Labor Department.
Over the past year, wholesale prices went up by 2.2% in May. When you don't count food
↪  and energy, core producer prices stayed the same from April to May but were up
↪  2.3% compared to last year. Economists watch the producer price index to
↪  understand how prices for consumers might change. Parts of this index are used in
↪  the Federal Reserve's personal consumption expenditures price index, which is
↪  their favorite way to measure inflation.
This news came after a report showed consumer inflation slowing down. Core consumer
↪  prices went up by just 0.2% from one month to the next, the smallest rise since
↪  October, and increased by 3.4% over the past year, the slowest in three years.
Inflation has dropped from a high of 9.1% two years ago, thanks in part to the Federal
↪  Reserve raising interest rates. Recently, the Federal Reserve decided not to
↪  change its main interest rate and said there might only be one rate cut this year.
```

```
Even though inflation is slowing down, everyday necessities like groceries, rent, and
↪  healthcare are still costly, creating political problems for President Biden.
↪  However, the U.S. economy is still strong, with low unemployment, steady hiring,
↪  and a better growth outlook from the World Bank, which now expects the economy to
↪  grow by 2.5% instead of 1.6%.
**Facts-only Title:**
U.S. Wholesale Prices Drop by 0.2% in May
**Facts-only Lead:**
U.S. wholesale prices fell by 0.2% from April to May, core producer prices remained
↪  stable month-to-month and consumer inflation growth hit its lowest level in three
↪  years.

--- SPECIFIC GUIDELINES ---
Craft a title and lead that reflect a facts-only perspective, taking the guidelines
↪  and examples above into account. Ensure the title is succinct, not longer than
↪  eleven words, and facts-only, and the lead provides a concise, accurate summary in
↪  a facts-only manner.
```

# List of Sources and Aids Used

The dissertation was prepared using the following sources and aids:

### Software and Technical Tools

- **Overleaf / LaTeX** for manuscript writing, formatting, and compilation.
- **Git** (version control) for project history and document management.
- **Statistical software** used for data preparation, estimation, and output generation:
  - Stata
  - R
  - Python

### Literature and Data Sources

- All scientific sources used are listed in the bibliography.
- Public and proprietary datasets used in the dissertation are described in the respective chapter sections and appendices.

### AI-Assisted Tools

- Generative AI tools (including large language models) were used as technical writing and coding assistants for tasks such as wording suggestions, editing support, and code debugging.
  - ChatGPT
  - Codex
  - Claude Code

No other aids beyond those listed above were used.

# Curriculum Vitae

## Education

**PhD Economics**, University of Cologne 2026

**MA Philosophy**, University College London 2020

**MSc Finance**, Stockholm School of Economics 2019

**BSc International Business**, Maastricht University 2017
Exchange semester: Tecnológico de Monterrey Campus Guadalajara

## Affiliations

**Member**, Center for Social and Economic Behavior (C-SEB) 2025 – Present

**Member**, Gateway Excellence Start-up Center Network 2024 – Present

**Associated Member**, ECONtribute: Markets & Public Policy 2021 – Present

## Academic Activities

*Research Visits*

ETH Zurich, Center for Law & Economics (Host: Elliott Ash), remote guest researcher 2026
Norwegian School of Economics, FAIR (Host: Ingar Haaland), 1 week 2026
Norwegian School of Economics, FAIR (Host: Ingar Haaland), 1 week 2025
Norwegian School of Economics, FAIR (Host: Ingar Haaland), 2 weeks 2024
International Monetary Fund, FIP (Host: Niamh Sheridan), 3 months 2023

*Conferences & Summer Schools*

Early-Career Behavioral Economics Conference (University of Chicago); 8th Economics of Media Bias Workshop (Frankfurt School of Finance & Management); 3rd Workshop in AI+Economics (ETH & University of Zurich); Economics of AI Conference (IESE Business School); ECONtribute Young Economist Workshop (University of Cologne); Bonn-Frankfurt-Mannheim PhD Conference (Goethe University); YEP Seminar (University of Cologne); IMPRS MPI Summer School (Max Planck Institute); Experimental Finance Summer School (University of Bonn).

*Refereeing*

Journal of the European Economic Association

*Departmental Service*

Co-organizer, ECONtribute Young Economist Workshop (annual) 2021 – 2023

## Work Experience

*Teaching*

**Lecturer**, Applied Econometrics (undergraduate), University of Cologne 2024 – Present
**Lecturer**, Survey Design (graduate), University of Cologne 2024 – Present
**Lecturer**, Experimental Methods (graduate), University of Cologne 2022 – Present
**Thesis Supervisor**, Economics (4× graduate), University of Cologne 2022 – Present
**Teaching Assistant**, Corporate Finance (graduate), Stockholm School of Economics 2018
**Tutor**, Finance I (undergraduate), Maastricht University 2015

*Pre-PhD Industry*

**Private Equity Analyst**, IK Partners (part-time) 2018 – 2019
**Consulting Summer Trainee**, Kearney 2018
**Finance Intern**, Rocket Internet 2017
**Controlling Intern**, Deutsche Telekom 2015

## Voluntary Work

**Student Ambassador**, Stockholm School of Economics 2018 – 2019
**Event Manager**, Study Association (SASSE) 2017 – 2019
**Treasurer**, Study Association (SCOPE) 2014 – 2015

# Affidavit

**according to Article 25 (5) of the doctoral degree regulations of 11 August 2025**

I hereby declare in lieu of oath that I have completed this dissertation independently and without the use of any aids or literature other than those specified.

No persons other than the co-authors listed in this dissertation were involved in its substantive preparation. All passages that are taken verbatim or are paraphrased from published and unpublished third-party works are labelled as such. I declare in lieu of oath that this dissertation has not yet been submitted to any other Faculty or university for examination; that it has not yet been published – apart from the stated partial publications and included articles and manuscripts – and that I will not publish the dissertation before completion of the PhD without the approval of the PhD committee. I am aware of the provisions of the PhD regulations of the PhD programme at the Faculty of Management, Economics and Social Sciences of the University of Cologne.

Furthermore, I hereby declare that I have read the guidelines of the University of Cologne for safeguarding good scientific practice and that I have observed them during execution of the work on which the dissertation is based and in the written dissertation itself. I hereby undertake to observe and implement the guidelines stated therein in all scientific activities. I confirm that the submitted electronic version corresponds completely to the submitted printed version. I assure you that I have told the truth to the best of my knowledge and have not concealed anything.

Cologne, 26 June 2026