WORKING PAPERS

N° 1519

March 2024

"Fake News: Susceptibility, Awareness and Solutions"

Tiziana Assenza, Alberto Cardaci, and Stefanie J. Huber

Toulouse School of Economics

Fake News:
Susceptibility, Awareness and Solutions

Tiziana Assenza[⋆] Alberto Cardaci[§] Stefanie J. Huber[‡]

Version: March 26, 2024

Abstract

This paper investigates and quantifies citizens' susceptibility to fake news and assesses, using a randomized control trial, the effectiveness of a policy intervention to raise awareness. We find that the average citizen lacks proficiency in identifying fake news and harbors an inflated perception of his/her ability to differentiate between true and fake news content. Increasing awareness by providing information about personal susceptibility to fall for fake news causally adjusts individuals' beliefs about their fake news detection ability. Most importantly, we show that the simple intervention of informing citizens about their personal susceptibility to fall for fake news causally increases their willingness to pay for the fact-checking service.

Keywords: Fake news, misinformation, disinformation, fact checking, information provision experiments, belief updating, willingness to pay.

JEL Classifications: C83, D83, D84, D91.

We are particularly grateful to Fabrice Collard, Thomas Dohmen, John Duffy, Patrick Fève, Sergei Guriev, Ingar Haaland, Emeric Henry, David Thesmar, Jean Tirole and Florian Zimmermann for precious comments. We also thank the participants of the 7^{th} Economics of Media Bias Workshop in Cologne for fruitful discussions. IRB approval was obtained from the Toulouse School of Economics. This experiment was registered at the AEA RCT Registry with ID AEARCTR-0007616. Tiziana Assenza acknowledges funding from the French National Research Agency (ANR) under the *"Investissements d'Avenir"* program, grant ANR-17-EURE-0010, and Stefanie Huber gratefully acknowledges support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2126/1-390838866.

⋆ Corresponding author. Toulouse School of Economics, University of Toulouse Capitole, IAST. E-mail: tiziana.assenza@tse-fr.eu. Phone: +33(0)5611 2882.

§ Goethe University, Frankfurt. E-mail: albertocardaci@gmail.com

‡ Department of Economics, University of Bonn; ECONtribute Research Cluster. E-mail: shuber@uni-bonn.de.

"You can avoid reality, but you cannot avoid the consequences of avoiding reality."

Ayan Rand, philosopher (1934–1982)

1 Introduction

In an address at the World Economic Forum in January 2024, Ursula von der Leyen, President of the European Commission, highlighted that *"For the global business community, the top concern for the next two years is not conflict or climate. It is disinformation and misinformation, followed closely by polarisation within our societies."* In this context, the need to understand how citizens navigate the information landscape is growing in importance. Knowledge of the complex relationship between individuals' ability to distinguish between actual and fake news, their beliefs about their susceptibility to fall for fake news, awareness, and economic behavior remains limited. Citizens might not only differ in their ability to detect fake news but also in their awareness of their susceptibility to fake news. Underestimating personal vulnerability might lead to a false sense of security and foster the penetration of misinformation. Beliefs about personal susceptibility and awareness can significantly affect economic decision-making.

We advance the literature by addressing the following unexplored question: Would citizens use information about their personal fake news detection ability and change their willingness to pay to protect themselves from misinformation via the channel of updated beliefs about their susceptibility to fall for fake news? This paper aims to fill this knowledge gap in the literature.

Recent surveys reveal that the average citizen is aware of the potential harms of fake news. 85% of European citizens believe fake news is a problem *"at least to some extent"* for Europe (Eurobarometer, 2018), similarly 82% of Americans believe that made-up news and information is a least a *"moderately big problem"* for the Unites States (Pew, 2019). Despite these expressed concerns and recognition of the potentially stark consequences of fake news, most citizens believe they are "immune" to being fooled by fake news. According to recent surveys, 84% of Americans and 71% of Europeans feel at least "somewhat confident" in their ability to detect fake news.[1]

Most of the related literature in psychology, political science, and economics has focused primarily on understanding the factors driving the consumption and sharing of fake news—with a

[1] For the US, this fraction varies between 75% and 93%, depending on the year when the survey was conducted. We refer to the Pew (2016a) Research Survey, the Statista (2020) Survey, and the YouGov (2023) Survey. For Europe, see the Flash Eurobarometer (2018) Survey.

strong focus on political fake news.[2] In addition, the fake news literature investigates how the prevalence of political fake news might have determined election outcomes in the past (Tworzecki, 2019; Enyedi and Kreko, 2018; Allcott and Gentzkow, 2017). The literature also investigates more general determinants of the sharing of fake news, such as information overload[3] or motivated beliefs—individuals wish to believe news that is in line with their ideology (Benabou and Tirole, 2011; Benabou, 2013; Zimmermann, 2020), and investigates the effectiveness of interventions to reduce the sharing of political fake news on social media (Guriev et al., 2023; Henry et al., 2022; Pennycook et al., 2021).

Recent literature takes a step back and investigates citizens' actual ability to detect fake news and associated socioeconomic differences. Angelucci and Prat (2024), Lyons et al. (2021) and Pennycook and Rand (2020a) focus exclusively on news related to politics, Arechar et al. (2023) on news related to Covid-19, while Arin et al. (2023) focus on general news. However, these studies have limitations. First, concentrating on one particular topic makes it difficult to disentangle citizen's knowledge about the specific topic from their ability to discern the truth.[4] Second, these studies are not incentivized. Finally, it is difficult to learn about the citizens' ability to distinguish between fake and true news if respondents know the probability distribution of fake/true news stories they must rate.

Our paper differs from these papers in at least four important aspects. First, we incentivize the choices of our survey participants. Second, we do not inform the participants about the probability distribution, as we are interested in understanding the absolute probabilities of truth individuals assign to the statements they are given to read. Third, we reduce the confounding between knowledge and ability by covering a representative spectrum of topics and news outlets. Fourth, and most importantly, the research question differs. This paper deciphers whether individuals hold accurate beliefs about their susceptibility to fall for fake news, using similar fake news detection ability quizzes. Most importantly, we study whether informing citizens about their actual susceptibility to fall for fake news causally shifts their beliefs and, in turn, affects their economic choices to protect themselves from misinformation.

[2]Political ideology is a strong motivation for sharing fake political news, which is driven mainly due to the feeling of hate toward political opponents (Osmundsen et al., 2021). An additional mechanism that encourages the spread of political misinformation is simply "group pressure" (Lawson et al., 2023).

[3]Humans have a limited attention capacity to digest information. For example, in the context of the Covid-19 pandemic, studies have shown that information overload increases the likelihood of sharing fake news by increasing the psychological strain of citizens (Apuke and Omar, 2021; Bermes, 2021). Similarly, Gleick (2011) and Angelucci and Prat (2024) argue that citizens "might be unable to acquire accurate information because (they) are drowning in an ocean of irrelevant and false information".

[4]According to the political science literature, men tend to be more likely than women to engage in partisan political participation (Quaranta and Dotti Sani, 2018). Women also tend to declare less interest in politics than men (Fraile and Gomez, 2017a,b). Hence, average women will be less knowledgeable about political topics. It will not surprise that women's "ability" to detect fake news is worse than men's.

Thus, the key contribution of this paper is to answers the critical unexplored research question of how raising awareness of fake news susceptibility at the individual level affects individuals' self-perceived ability to detect such news and subsequent economic decision-making. To this end, we designed a survey with an RCT information provision experiment in order to exogenously shift individuals' beliefs about their perceived susceptibility to fall for fake news. Subsequently, we test whether reducing individuals' misapprehension of their susceptibility affects their willingness to act against the risk of being harmed by misinformation. Our design advances the related literature using information provision experiments—as we investigate the effect of (changes in) beliefs on behavior. Hence, the effect of the information treatment on beliefs serves as a manipulation check. The goal of the experiment is to assess the causal effect of beliefs on behavior, and in order to assess this, we exogenously change beliefs using an information treatment. The manipulation check (of changes in beliefs) serves the purpose of documenting that the exogenous shift was successfully induced.

Our experiment has several attractive features: First, it was carefully designed to be simple and engaging (as described in detail in Stantcheva, 2023). Attention checks and participants' post-experiment feedback indicate that levels of engagement were in fact high. Moreover, we strove to reduce the potential for experimenter demand effects by choosing a between-subject instead of a within-subject design (e.g., Charness et al., 2012 and Zizzo, 2010). In addition, we used three key characteristics recognized as effective countermeasures in the literature on information provision survey experiments to mitigate the experimenter demand effects i.e., anonymity, incentivized tasks, and neutral framing (Haaland et al., 2023).

To make participants aware of their susceptibility to fake news, they performed an incentivized news accuracy detection task in which they had to distinguish between accurate and fake news. We asked them to assess the accuracy of the information content in a series of 20 news headlines displayed in sequence, randomized at the individual level.[5] Each headline contained a subheading highlighting the critical content of the news and the publication date. Importantly, our focus is exclusively on the accuracy of the information based on the news content, not its visual aspects (e.g., fabricated pictures, grammatical errors). For this reason, all our news items share the same appearance, with only text elements but no visual components, such as pictures. We elicited the respondents' quantitative beliefs about their ability at three moments during the fake news

[5]We collect twenty news headlines from a set of fact-checked news items by different non-partisan organizations (including Politifact, Snopes, Reuters, Science Feedback, and Factcheck.org). All selected news items appeared in the news media between May 2018 and May 2021. Half of these news items are classified as containing accurate information, i.e., their primary elements are correct and demonstrably true. The second half is classified as inaccurate information i.e., their primary elements are incorrect and demonstrably false. We do not provide information about this distribution but inform participants that independent third parties have fact-checked all news items.

detection task. First, after explaining the task and before starting the first part of the task, we measured prior beliefs. Second, we elicited their mid-task beliefs after exposing subjects to the first ten news items. Finally, after completion of the task, we elicited posterior beliefs.

We randomly assign 50% of the subjects to our information treatment—that reveals participants' personal performance to detect fake news, our proxy to measure individual susceptibility to fall for fake news. Treated subjects receive information on their actual susceptibility after the first part of the task; treatedsubjects learn how many of the ten news items they correctly identified as accurate or inaccurate news. Information provision occurs after eliciting their mid-task belief and before starting the second part of the task. Therefore, we can analyze whether the information treatment successfully raised awareness of susceptibility to falling for fake news, resulting in changed beliefs.

Next, we investigate the treatment effect of providing information about subjects' susceptibility to fall for fake news on their economic choices—via the channel of beliefs. Specifically, we provide subjects (pre and post-experiment) with a hypothetical $1000 budget that they can freely allocate between consumption goods, health insurance, or misinformation insurance—an insurance contract that covers the risk of being harmed by inaccurate news and information (i.e. a service such as fact-checking). Hence, we can establish a causal relationship between changes in subjects' willingness to pay for misinformation insurance and the information provision about their personal susceptibility to fall for fake news. This design allows us to analyze the effect of the treatment on spending in relative terms. We choose the label "misinformation insurance" for two key reasons. First, the framing "misinformation insurance" is clearly more neutral than "fact-checking service". Second, it captures the notion of probability. Insurance reduces harm but cannot protect against bad outcomes with a 100% probability.

The key results of this paper can be summarized as follows. First, we find that the information treatment causes a significant change in subjects' beliefs about their ability to detect fake news. This change in awareness of their susceptibility to fake news (i.e. belief updating) leads to a more accurate self-assessment of personal susceptibility. Second, treated subjects who received a negative signal about their performance in the fake news detection task adjusted their ability beliefs downward. This shift in beliefs, induced by the information treatment, in turn, causally increases subjects' willingness to pay to reduce the risk of being harmed by misinformation.

Regarding socioeconomic characteristics, the academic literature lacks consensus on which characteristics predict the ability to detect fake news (Arin et al., 2023). Therefore, our objective is to provide systematic evidence on the effect of gender, age, and other individual-level characteristics on the ability to detect fake news. The news items were carefully chosen to allow assessment of the average population's ability to detect fake news, individuals' beliefs about their own ability, and the heterogeneity among socioeconomic characteristics. Instead of focusing on, e.g., exclusively political fake news (as in, e.g., Angelucci and Prat, 2024; Guriev et al., 2023; Pennycook and Rand, 2020a), statements of a single politician (as in e.g., Barrera et al., 2020) or a specific topic (as in e.g., Arechar et al., 2023; Lutzke et al., 2019), we aim to cover a representative spectrum of topics and news outlets. We choose heterogeneous news items along two essential dimensions; the publisher and the topic.[6]

We find that women, older respondents (older than 54 years), and those who are single perform significantly better in detecting fake news. Educational attainment and household income do not predict proficiency in detecting fake news. We find that better news accuracy detection performance correlates strongly with a higher score on the Cognitive Reflection Test (as in Bago et al., 2020 and Pennycook and Rand, 2019) and the Sentence Comprehension Test. The results are highly robust to adding information about subjects' media consumption habits or political preferences. A strong predictor is social media consumption; individuals that indicate that this is their primary source of information perform significantly worse.[7]

Our results are relevant from a policy perspective. The policy debate has considered a variety of supply and demand side interventions to address the problem of fake news. Supply-side solutions are oriented toward diminishing the amount of misinformation or disinformation supplied into the market. Measures in this regard can be privately driven, e.g., via social media platforms' policies on content moderation, or state-led, e.g., via regulations that dictate to commercial information providers (such as social media platforms) certain obligations on fact-checking. Some researchers have already tested the effectiveness of tools in this direction, including fact-checking, requiring confirmation clicks to access or share false content, or requiring a comment before sharing (e.g., Pennycook and Rand, 2020b; Henry et al., 2022; Arechar et al., 2023; Guriev et al., 2023).

While the supply-side approach has the advantage of addressing the problem at its roots, it comes with at least two disadvantages. First, the technology to create fake news develops fast. Those implementing supply-side policies—be those private or public actors—face challenges to keep up with the technological changes being driven by those wishing to spread fake news. Second, these supply-side solutions risk being politicized, e.g., right-wing politicians accuse social media

[6] Half of our news items appeared on mainstream news sources (e.g., FOX News, the Wall Street Journal, the New York Times), and the other half on alternative news sources (e.g., Naturalnews.com, Breitbart, Raw Story). In addition, we select news items covering a wide range of topics, such as crime, science, politics, the Covid-19 pandemic, climate change, and health.

[7] Our finding on the age dimension aligns with Pennycook and Rand (2019), Arin et al. (2023), and Angelucci and Prat (2024). Unlike Angelucci and Prat (2024), we find that women perform better than men in detecting fake news. This difference is most likely explained by the fact that Angelucci and Prat (2024) focus on political news; and men are just more engaged, interested, and informed about politics.

platforms of colluding with left-wing politicians and academic experts to censor the "truth" in the United States. As Stiglitz and Kosenko (2024) put it, "a central question in the design of such regulations is how to prevent such harms (created by fake news) within democratic frameworks that emphasize, for instance, freedom of speech (First Amendment rights)".

Demand-side policies would instead focus on influencing citizens' ability to detect fake news. Potential policies addressing the demand side should include improving general educational outcomes and specific training in media and digital literacy. These interventions would likely improve citizens' fake news detection performance but require financial investments and planning horizons.

This paper makes use of citizens' misperception of their personal susceptibility to fall for fake news. It proposes a simple and cost-effective tool—at least for those who trust fact-checking companies in the first place. We implement a policy intervention that educates people about their personal susceptibility to fall for fake news. Those who receive a negative signal about their fake news detection performance significantly adjust their ability-beliefs—and most importantly, take action accordingly to protect themselves from harms created by fake news.

The remainder of the paper is structured as follows. Section 2 describes the data. Section 3 discusses the design of the study and the implementation of the RCT information experiment. Section 4 presents descriptive statistics and our results on households' susceptibility and their awareness of their ability to detect fake news. In addition, this Section examines heterogeneity among different socio-economic groups. Section 5 presents our key results of the RCT information provision experiment. Section 6 discusses the policy implications of our findings and concludes.

2 Data

2.1 News Headlines Data

In this paper, we define *fake news* as demonstrably false information (as in Acemoglu et al., 2023 and Allcott and Gentzkow, 2017). For our study, it is not necessary to differentiate between disinformation (intentionally misleading) and misinformation (not intentionally misleading). We built our news headlines dataset by selecting twenty news headlines from a set of fact-checked news items. All news items have been fact-checked by at least one of the non-partisan organizations (including Politifact, Snopes, Reuters, Science Feedback, and Factcheck.org); some of the news items have been fact-checked by more than one organization. All selected news items appeared in the news media between May 2018 and May 2021. Half of these news items are classified as containing accurate information, i.e., their primary elements are correct and demonstrably true. The second half is classified as inaccurate information i.e., their primary elements are incorrect and demonstrably false. We do not provide information about this distribution but inform participants that independent third parties have fact-checked all news items.

To assess individuals' general susceptibility to fall for fake news, we carefully choose the news items. Instead of focusing on, e.g., statements of a single politician (as in e.g., Barrera et al., 2020) or specific topic (as in e.g., Lutzke et al., 2019 on fake news about climate change on Facebook or Arechar et al. (2023) on the Covid-19 pandemic), we aim to cover a representative spectrum of topics and news outlets. Therefore, we choose heterogeneous news items along two essential dimensions; the publisher and the topic. Half of our news items appeared on mainstream news sources (e.g., FOX News, the Wall Street Journal, the New York Times), and the other half on alternative news sources (e.g., Naturalnews.com, Breitbart, Raw Story). In addition, we select news items covering a wide range of topics, such as crime, science, politics, the Covid-19 pandemic, climate change, and health.

2.2 The Survey: Individual Data and Representativeness

This paper uses a novel large-scale survey data set ($N = 2,413$) collected between June 11 and August 6, 2021.[8] The company Qualtrics collected the data on our behalf using non-interlocking quotas to ensure the sample's representativeness along the dimensions of age, gender, ethnicity, and region within the United States. We restricted the sample to respondents born in the United States and 18 years or older. The representativeness of our sample is investigated in detail in the Appendix. Appendix Figure A1 reports the distribution of the sample for the key socioeconomic characteristics and Appendix Table A1 shows that the sample is representative of the general US-born population (aged 18-year-old and older) on gender, age, education, region of residence, marital status, household size, share of Black/African Americans, party affiliation, and, to a lesser extent, in income.

Compared to most laboratory experiments, one major advantage of our online survey experiment is its external validity, thanks to the large sample size and the data' representativeness of the US population. However, one potential disadvantage is related to internal validity considerations. Participants may pay less attention when participating online compared to in-person laboratory experiments. To address this concern, we designed the survey (experiment) with the objective to be engaging for the participants, and implement a simple and widely used attention check (see, e.g., Faia et al., 2022; Roth and Wohlfart, 2020) to screen out distracted participants leading to

[8] After clicking on the survey link, participants receive a consent form providing information about the nature and research purposes of the survey. In particular, they are informed that they are participating in an academic research survey and that participation is anonymous and voluntary. The median survey completion time was 20 minutes and 34 seconds.

potentially low-quality observations.[9] 75% of all respondents passed the attention check and are included in this study.[10] Figure 6 in Section 5 shows that the participants who completed the survey were engaged and found the survey interesting.

In addition to the socioeconomic characteristics discussed above, we elicit a wide range of attitudes, beliefs, and trust in news and information. Appendix Tables A2–A3 report the descriptive statistics of these variables. To measure the trust in and consumption of news and information, we include a rich set of questions about subjects' level of trust in others, trust in non-partisan fact-checkers, their news consumption habits, and primary news sources (television, newspapers, social media, etc.)—including the level of trust in each of them. We also add a question to understand whether and to what extent subjects consider misinformation a problem for the country.

In addition, we elicit respondents' risk aversion, text comprehension skills, and tendency to rely on intuitive thinking—as these variables could in principle determine the ability to detect fake news. Appendix Table A4 reports the descriptive statistics of these variables. To elicit respondents' risk aversion, we follow Dohmen et al. (2011) by asking about their general willingness to take risks on a scale from 0 to 10, where 0 indicates that they are not willing to take risks and 10 indicates intense risk-taking. To measure subjects' ability to comprehend sentences with varying levels of syntactic complexity, we use the test developed by Vernice et al. (2019). We use this variable later in the analysis to investigate whether the ability to assess the accuracy of news relates to subjects' text comprehension skills. Bago et al. (2020) showed that greater deliberation, measured by higher scores in the Cognitive Reflection Test (CRT), predicts a greater ability to distinguish the accuracy of news items. Therefore, we elicit subjects' tendency to rely on intuitive thinking via the CRT (Frederick, 2005) to control for this ability in our analysis.

3 Experimental Design

3.1 Structure of the Survey Experiment and Hypotheses

This section describes the experimental (between-subject) design for the empirical analysis in Sections 4 and 5. Our goal is threefold. First, we aim to quantify individual's susceptibility to fall for fake news and their beliefs. Second, our objective is to test, using an RCT information provision experiment, the effectiveness of a simple policy intervention to increase awareness of susceptibil-

[9] The attention check asks the following question *"To show that you read our questions carefully, please choose "Very Strongly interested" and "Not at all interested" as your answer to this question: How interested are you in politics?"*. They are given five scale options from "Very Strongly interested" and "Not at all interested".

[10] As a comparison, in Chopra et al. (2022), only 56% of the respondents passed the attention check, which is very low compared to many other experiments (e.g., 96.4% in Bottan and Perez-Truglia (2022) and 99% in Nathan et al. (2020)).

ity to fall for fake news at the individual level. Third, we investigate the impact of this simple policy intervention—informing participants about their actual ability to distinguish between true and fabricated information—on their willingness to pay for misinformation insurance. We test the following two hypotheses.

Hypothesis H1 (Ability belief updating): *Treated subjects (conditional on trusting the signal) update their belief and achieve a more accurate self-assessment of their susceptibility to fake news.*
Hypothesis H2 (Economic choices): *Subjects receiving a negative signal, conditional on H1 being true, significantly increase their willingness to pay for misinformation insurance.*

Hypothesis H1 examines whether the information treatment causes a significant change in subjects' belief updating and accuracy. Hypothesis H2, assuming Hypothesis H1 is true, investigates whether the treatment causally increases the willingness to pay for misinformation insurance among subjects receiving negative feedback i.e., individuals with an inflated perception of their fake news detection ability prior to receiving feedback on their actual ability.

Figure 1 illustrates the timeline and basic structure of the information provision experiment, which consists of three essential parts.

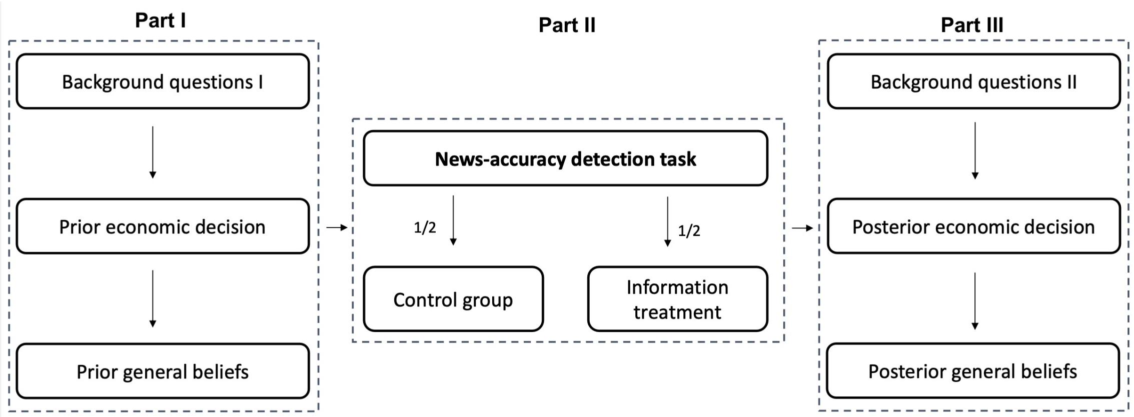
- **Part I (Individual's prior general beliefs and behavior):** Eliciting subjects' qualitative general beliefs about their own and the average American's ability to detect the accuracy of news and information. Measuring subjects' economic choices pre-experiment. Subjects must allocate their budget between consumption goods, a health insurance, and a misinformation insurance (see Section 3.2).

- **Part II (Experiment/Treatment):** All subjects perform the incentivized "news-accuracy detection" task. For all subjects, we elicit quantitative beliefs during three moments of the task. Subjects in the treatment group receive information related to their ability to detect the accuracy of news and information, after the first and before the second part of the task (Figure 3). In the case of the control group without information provision, there is no signal. The task, the information provided, and the elicited quantitative beliefs are described in Section 3.3.

- **Part III (Individual's posterior general beliefs and behavior):** Eliciting subjects' qualitative general beliefs about their own and the average American's ability to detect the accuracy of news and information. Measuring subjects' economic choices post-experiment. Subjects must allocate their budget between consumption goods, a health insurance, and a misinformation insurance (see Section 3.2).

The main analysis consists of measuring how the information provided to individuals changes their awareness about their own and the average American's susceptibility to fall for fake news and how this information changes their beliefs and their economic choices.

Figure 1: Survey Structure and Timeline



3.2 Measuring Individual's General Beliefs and Behavior

General qualitative beliefs. We elicit respondents' general qualitative beliefs about their own and the average American's ability to detect the accuracy of news items. Subjects can answer on a scale from 1 (very bad) to 5 (very good) for both questions. We ask these questions before and after the news-accuracy detection task, thus eliciting prior and posterior beliefs. Specifically, we ask subjects the following two questions:

1. *"In your opinion, how good is your ability to identify news or information that misrepresents reality or is even false?"*

2. *"In your opinion, how good is the average American's ability to identify news or information that misrepresents reality or is even false?"*

Economic choices. In addition to changes in beliefs, we also investigated the treatment effect of providing information about subjects' news-accuracy detection ability on their actual economic choices. Specifically, we provide subjects (pre- and post-experiment) with a hypothetical $1000 budget that they can freely allocate between consumption goods, health insurance, or *misinformation insurance*—an insurance contract that covers against the risk of being harmed by inaccurate news and information (i.e., a service such as a fact-checker).

This design allows the analysis of the effect of the treatment on spending in *relative* terms. The label "misinformation insurance", we choose for two key reasons. First, the general public understands the concept of "insurance", which captures the notion of probability. An insurance reduces the (potential) harm and cost of bad outcomes realized, but cannot protect against it with probability 100% .[11] Second, the notion of a "fact-checker service" could polarize. The framing "misinformation insurance" is clearly more neutral than "fact-checker service". Therefore, we opted for this framing. Hence, using the notion of "misinformation insurance" allows us to reduce confounding factors while measuring the demand to be protected from misinformation.

Therefore, we can establish a causal relationship between changes in subjects' willingness to pay for misinformation insurance and the provision of information about their news accuracy detection performance. In other words, we measure the treatment effect on the economic choices by eliciting quantitative variations in spending in a "hypothetical" scenario (as in e.g., Fuster et al. (2021); Parker and Souleles (2019); Christelis et al. (2019); Jappelli and Pistaferri (2014)). This reported preference approach is widely used in the related literature as it provides great flexibility in treatment design (Fuster et al., 2021). In the field of economics, Parker and Souleles (2019) and Parker et al. (2013) study the difference between the reported (survey) data and the revealed consumption expenditures. This research shows that self-reported data works well to predict behavioral changes and estimate population aggregates, which is the objective of this study.[12]

3.3 Information Treatment and News-Accuracy Detection Task

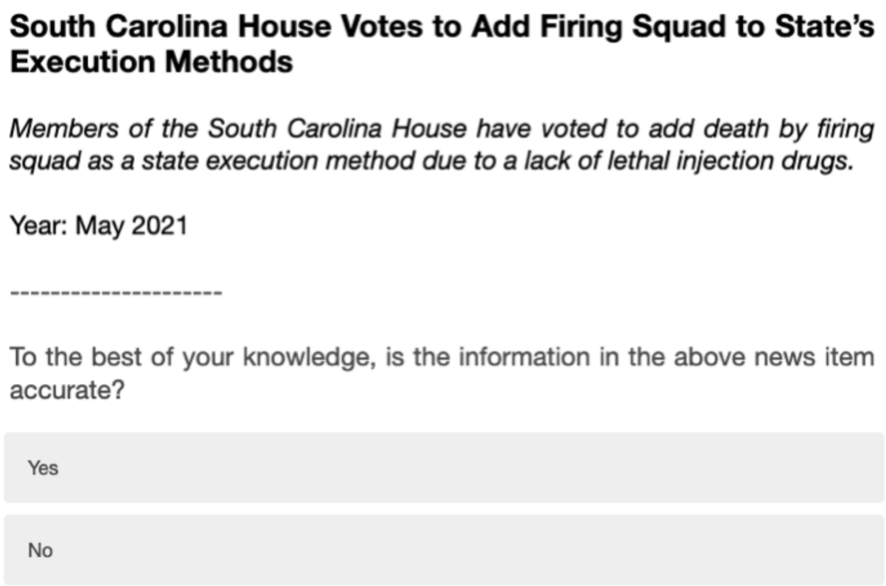
News Accuracy Detection Task. Participants are asked to assess the accuracy of the information content in a series of 20 news headlines displayed in sequence, randomized at the individual level. Half of the news headlines provide accurate information, and the other half is fake news. We do not provide information on this distribution, as we are interested in the absolute probabilities

[11] A health insurance cannot protect you from getting cancer, but can reduce the probability of death and the costs associated with medical treatments. Similarly, a misinformation insurance (a service like a fact-checker) can reduce the risks of being harmed by fake news. A fact-checking service provides accurate information most of the time, but not always. Mistakes may occur and will be corrected over time.

[12] In general, the literature finds different effect sizes for hypothetical and non-hypothetical choices. For example, the literature highlights stronger effects on voting intention than on actual voting (Chiang and Knight (2011); Gerber et al. (2011); Gerber et al. (2009)). However, for consumption decisions there is less of a difference. E.g., Parker and Souleles (2019) compares reported consumption responses to hypothetical tax rebates with the actual spending responses of past tax rebates. They find a negligible difference. Similarly, Parker et al. (2013) find that the reported preferences match the actual behavior. Subjects who reported spending their 2008 fiscal stimulus payment are found to have done so. Similarly, in the field of decision neuroscience, Kang et al. (2011) used a simple functional magnetic resonance imaging (fMRI) task in which subjects make real and hypothetical purchase decisions for consumer goods. They found that the same types of valuation and choice computations are performed in hypothetical and real consumption decisions.

of truth individuals assign to the news items.[13] Each headline also contains a subheading highlighting the critical content of the news and the publication date. Figure 2 shows how the news items appear to the subjects. Importantly, our focus is exclusively on people's ability to discern the accuracy of the information based on the news content, not its visual aspects (e.g., fabricated pictures, grammatical errors). For this reason, all our news items share the same appearance, with only text elements but no visual components, such as pictures.[14]

Figure 2: Appearance of news items in the news-accuracy detection task


South Carolina House Votes to Add Firing Squad to State's Execution Methods

Members of the South Carolina House have voted to add death by firing squad as a state execution method due to a lack of lethal injection drugs.

Year: May 2021

To the best of your knowledge, is the information in the above news item accurate?

Yes

No

As shown in Figure 2, we ask participants to provide a binary true / false rate on the headlines of the news presented (e.g., as in Bago et al., 2020 and Ross et al., 2021). Unlike previous studies by Angelucci and Prat (2024) and Henry et al. (2022), we deliberately forgo the option *"I do not know"* or a probability rating. We want the participants to make a clear choice. Recall that we use the accuracy detection task to inform participants of their actual ability to discern accurate and fabricated information. This feedback mechanism served to raise awareness of their individual susceptibility to fall for fake news. The core objective of our paper is to investigate whether increased awareness of personal susceptibility to fall for fake news impacts the demand for protection from misinformation.

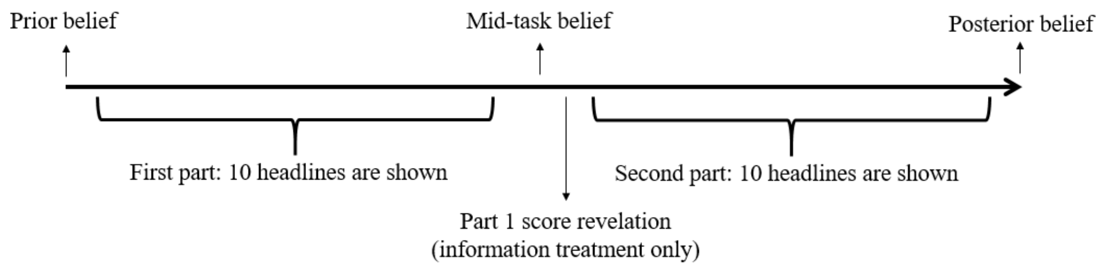
[13] This approach differs from Angelucci and Prat (2024), who study relative probabilities by showing six news items on the same screen and informing the participants that 50% of them are true.

[14] Results of a treatment variation in which the presentation of the news items additionally contains the news source are available on request. In brief, displaying the news source does not affect the key findings presented in this paper.

Incentives. To incentivize the task, participants receive one point for each correct answer to the following question: *"To the best of your knowledge, is the information in the above news item accurate?"* (see Figure 2). Recall that participants will answer this question after each of the 20 news headlines they will see throughout the experiment. The total number of points is converted into $US at a conversion rate of 1 point = $5, and their converted payoff might be donated to a charity of the participant's choice.[15] At the end of the experiment, two participants will be randomly selected and their total scores will form the basis of the donation.

Elicited quantitative beliefs during the task. Figure 3 shows that the news-accuracy detection task has two parts; each consists of ten news items randomly selected from the 20 headlines. We elicit respondents' quantitative beliefs in three moments during the task (Figure 3). First, after explaining the task and before starting the first part, we elicit *prior beliefs* by asking the following question: *"How many points do you think you will score?"*. We instruct subjects to provide a value between 0 and 10. Secondly, after exposing subjects to the first ten news items, we elicit their *mid-task beliefs* through the following question: *"You have seen ten news items. How many points do you think you scored?"*. Eventually, subjects begin the second part of the task and assess the accuracy of the remaining ten news items. Finally, after the second part of the task, we elicit *posterior beliefs* by asking subjects to answer this question: *"You have seen another set of 10 news items. How many points do you think you scored?"*. Moreover, for each of the three above questions, we also elicit subjects' confidence in their beliefs by asking them how sure they are about their answers, from *"Very unsure"* to *"Very sure"*.

Figure 3: Timeline of quantitative beliefs and news-accuracy detection task



Information treatment. We randomly allocate 50% of the subjects to our *score revelation treatment*. Treated subjects receive information on their score after the first part of the task (i.e.,

[15] Hence, the individual earnings range from zero to twenty points (e.g., zero to $100). During the survey, we asked participants whether they preferred the Feeding America or American Red Cross charity.

subjects learn how many of the ten news items they correctly identified as accurate or inaccurate news). The information provision occurs after eliciting their mid-task beliefs and before starting the second part of the task (see Figure 3).

Potential Design Concerns. We reduced the potential for experimenter demand effects by choosing a between subject design compared to a within subject design (e.g., Charness et al. (2012) and Zizzo (2010)). In addition, our experimental design is characterized by three central features that the literature considers as powerful tools against experimenter demand effects when conducting information provision survey experiments, namely anonymity, incentivized tasks, and neutral framing (see Haaland et al. (2023) for a detailed discussion).

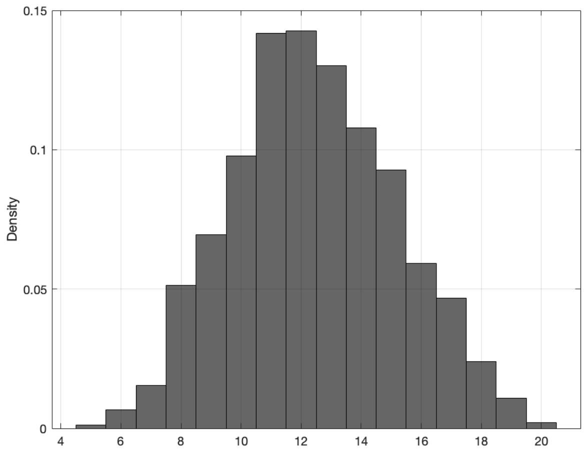
4 Descriptive Statistics and Analysis

First, this section provides descriptive statistics on the average American's awareness and susceptibility to fall for fake news. Throughout the reminder of the paper, we use the terms susceptibility to fall for fake news and ability to detect fake news interchangeably. Second, this section investigates whether specific socioeconomic characteristics predict the ability to detect fake news. Regarding socioeconomic characteristics, the academic literature lacks agreement on the characteristics that predict the ability to detect fake news (Arin et al., 2023). Therefore, our objective is to provide systematic evidence for the US population on the effect of gender, age, and other individual-level characteristics such as marital status, education, employment status, or income on the ability to detect fake news. In contrast to most of the related literature, this paper reduces the confounding between knowledge and ability by covering a heterogeneous spectrum of topics and news outlets.

To determine the participants' susceptibility to fall for fake news, we let them assess the accuracy of twenty news items in the news-accuracy detection task (see Figure 2). For each correct answer, one point is received. Therefore, if the participant perfectly distinguishes between accurate and fake news, the total score equals 20 points. On the contrary, 20 random draws would lead to an average score of 10 points. Figure 4 shows the total score distribution of the sample. The average score equals to 12.45 out of 20, i.e., respondents rated 62.25% of the news items correctly. This result is strikingly in line with Maertens et al. (2023), who show that – on average – adult US citizens correctly classified two-thirds (65%) of the headlines that they were shown as either real or fake. This finding shows that respondents' ability to differentiate between accurate and inaccurate news items is far from perfect, and thus, the problem of fake news is substantial.[16]

[16]Interestingly, participants seem equally likely to *fall for inaccurate news* and to *mistakenly believe that accurate news is false.* Participants gave the correct assessment in 62.55% of the accurate news and in

Figure 4: Distribution of total score in news-accuracy detection task



Notes: Participants assess the accuracy of twenty news items in the news accuracy detection task by answering the question: *"To the best of your knowledge, is this information in the above news item accurate?"*. The answer categories are *"Yes"* or *"No"*. For each correct answer, one point is received. This figure shows the total score (point) distribution for the entire sample; with an average score of 12.45.

Next, we analyze subjects' general qualitative beliefs about their own and the average American's ability to distinguish between accurate and inaccurate information. The large majority has tremendous confidence in their ability to evaluate news accuracy; 82.64% report having a *"good"* or *"very good"* ability to identify news or information that misrepresents reality or is even false.[17] Interestingly, only 37.61% of our respondents believe that the average American can identify news or information that misrepresents reality or is even false—providing evidence of overplacement à la Moore and Healy (2008).[18]

Result 1 (Awareness and Susceptibility to fall for Fake News.)**.** *The average American believes s/he is "good" or "very good" at distinguishing between fake and accurate news. However, the average American performs poorly and is susceptible to fall for fake news.*

61.99% of the inaccurate news items. Furthermore, we could not detect any time trends. There is no significant difference between the average score in the first and second part of the task ($p = 0.301$).

[17] This is strikingly in line with recent Pew Research Survey findings, e.g., Pew (2016a) and Pew (2016b), where 80% of the respondents say this statement describes them *"very well"* (41%) or *"somewhat well"* (40%): *"Most of the time, it is easy for me to determine what information is trustworthy."*

[18] This result is consistent with the psychology literature finding that individuals not only tend to have positive self-perceptions, but that they often believe they are more talented and competent than others, even when they are not (e.g., Anderson et al., 2012; for reviews, see Alicke and Govorun, 2005; Dunning et al., 2004). Instead of these general and qualitative beliefs, we use for the primary (treatment effect) analysis subjects' quantitative beliefs about their ability to distinguish between accurate and inaccurate information. These quantitative beliefs were elicited three times during the news-accuracy detection task (Figure 3). In particular, we focus on the mid-task beliefs, allowing subjects to familiarize themselves with the task first. The results of this paper are robust to using the prior instead of the mid-task beliefs. Results are available on request.

As Figure 4 illustrates, the ability to distinguish between accurate and inaccurate information varies considerably between participants. We identify substantial heterogeneity in the ability to assess the veracity of information in the news headlines among participants. A considerable share (24.2%) provides the correct assessment for less than half of the news items, doing less well than a random draw. Only 14.29% of the participants correctly assessed the veracity of information for at least 80% of the news headlines.

To understand this heterogeneity, we first analyze whether certain socioeconomic or behavioral factors correlate with the ability to distinguish between accurate and inaccurate information. We estimate the following baseline model using OLS (ordinary least squares):

$$SCORE_i = \beta_0 + \beta_1' \boldsymbol{X_i} + \beta_2' \boldsymbol{B_i} + \beta_3' \boldsymbol{I_i} + \varepsilon_i, \tag{4.1}$$

where $SCORE_i \in [0, 20]$ denotes the total score of news-accuracy detection task. X_i denotes a vector of socioeconomic controls for the individual i, including gender, age, marital status, household income, educational attainment, employment status, categories for ethnicity (race), location of residence, a measure for cognitive ability, the cognitive reflection test (CRT) score, and a measure for text comprehension ability, the Sentence Comprehension Test (SCT) score. $B_{i,t}$ denotes a vector of behavioral factors and experiences, including the individual's i perceived trust in fact checking, general trust in others, the individual's beliefs i on whether misinformation is a problem for the country, and the individual's political preferences i. The vector I_i includes several measures of the individual's i reported news consumption. The error term is denoted by $\varepsilon_{i,t}$.

Table 1 shows the main results of the estimation of the model in (4.1).[19] Column 1 shows the estimation results using exogenous controls for the individuals, namely, the female dummy, age, and race categories. In addition, we include the test results for the Cognitive Reflection Test (CRT) and the Sentence Comprehension Test (SCT)—two measures consistently predicting truth discernment in the related literature.[20] In column 2, we add the remaining variables of the vector X_i, the endogenous socioeconomic controls, namely, marital status, employment dummy, categories for education and income, and type of location. In column 3, we add the behavioral controls B_i. In column 4, we add the media consumption proxies I_i. Finally, in Column 5, we show our main results as specified in our baseline specification (4.1).

For all specifications, we find that a higher news accuracy detection score correlates strongly with a higher score in the Cognitive Reflection Test and the Sentence Comprehension Test. This result aligns with existing evidence on the importance of analytical and deliberate thinking for

[19]Table 1 only reports significant results. Appendix Table A6 shows the full estimation results.
[20]Excluding the CRT and SCT test scores do not change the result in Column 1.

Dependent variable: Total Score of the news-accuracy detection task					
	(1)	(2)	(3)	(4)	(5)
Female	0.277**	0.226*	0.307***	0.295***	0.212*
	(0.11)	(0.12)	(0.11)	(0.11)	(0.11)
old	1.321***	1.169***	1.468***	1.086***	1.069***
	(0.14)	(0.16)	(0.14)	(0.15)	(0.17)
score sct	0.661***	0.590***	0.642***	0.623***	0.541***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
score crt	0.473***	0.478***	0.433***	0.447***	0.422***
	(0.07)	(0.07)	(0.06)	(0.07)	(0.06)
single		0.554***			0.550***
		(0.12)			(0.12)
currently employed		-0.409***			-0.456***
		(0.13)			(0.13)
living in suburbs		0.401***			0.506***
		(0.13)			(0.13)
living in countryside		0.166			0.335**
		(0.17)			(0.17)
Behavioral factors B					
trust in fact checking			0.505***		0.535***
			(0.11)		(0.12)
political preference: Independent			-0.059		-0.247*
			(0.13)		(0.14)
political preference: Prefer not to answer			-0.530**		-0.695***
			(0.23)		(0.24)
political preference: Republican			-0.928***		-1.029***
			(0.13)		(0.14)
Media Consumption I					
Media Consumption Indicator				0.095**	0.072
				(0.05)	(0.05)
social media consumption				-0.858***	-0.713***
				(0.15)	(0.16)
Constant	Yes	Yes	Yes	Yes	Yes
Endogenous Controls	No	Yes	No	No	Yes
Observations	2324	2224	2324	2324	2224
R^2	0.16	0.18	0.19	0.17	0.23

Notes: Columns 1-5 report OLS estimates. Robust standard errors (Eicker-Huber-White) are reported in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Exogenous socio-demographic controls include gender, age (reference category: young, middle aged, old), and categories for ethnicity (Non-Hispanic White, Hispanic, Black or African American, American Indian or Alaska, Asian, Native Hawaiian or Pacific Isander, Other). SCT denotes the score of the Sentence Comprehension Test. CRT denotes the score of the Cognitive Reflection Test. Additional (endogenous) socio-demographic controls in Column (2) and (5) include marital status, household income (reference category: low, middle, high), education (reference category: low, middle, high), employment dummy, location of residence type (suburb, countryside, omitted category: living in city). The behavioral variables B include individual's trust in fact checking (5-point answer scale: strongly distrust to strongly trust), general trust in people (5-point answer scale: strongly distrust to strongly trust), perception of fake news being a problem for the country (4-point answer scale: very big problem to not a problem at all), and political preference (reference category: democratic, republican, independent, prefer not to answer). Media consumption I variables include the media consumption indicator and measures for the consumption frequency of four different news sources: social networks, television, online newspapers, printed newspapers; for each (5-point answer scale, not familiar with this source of news to main source of news). The dependent variable (total score) is measured by the news-accuracy detection task. The participants assess the accuracy of 20 news items in the news accuracy detection task by answering the question *"To the best of your knowledge, is this information in the above news item accurate?"*. The answer categories are *"Yes"* or *"No"*. For each correct answer, one point is received.

Table 1: Ability to detect Fake News (Socioeconomic and behavioral characteristics)

truth discernment (Bago et al., 2020; Pennycook and Rand, 2019). Although we cannot claim causality, our finding suggests that effort in thinking and the ability to understand the information content of a text are significant predictors of the capacity to assess the accuracy of information.

We find a positive and statistically significant impact of the female dummy for all specifications. Furthermore, we find a consistent and statistically significant impact of age, marital status, employment dummy, and location residence type for all specifications. Older respondents (older than 54 years), those who are single, those who are not employed, and those who live in a suburb or countryside (and not in a city) perform significantly better at detecting fake news. Educational attainment and household income do not predict the ability to detect news.[21]

Next, we turn to the behavioral factors B_i. In columns 3 and 5, we include individuals' trust in fact-checking, general trust (in people), their tendency to consider misinformation a problem for the country, and their political preferences. We find that trust in fact-checking positively correlates with detecting fake news, while general trust in people does not. Interestingly, political preferences matter. Republicans (and Independents, to a lesser extent) perform significantly worse than Democrats in detecting fake news. Our last finding concerns individuals' media consumption I_i. Column 4 shows that those consuming information from a more extensive set of sources – including television and radio, printed and online newspapers, and social networks – have a better assessment of the accuracy of the news items.[22] However, this result does not hold for the full baseline specification (column 5). In contrast, for all specifications, we find that those who use social media as a primary source of news consumption were the least able to distinguish between inaccurate and accurate news items.

The few existing papers investigating the gender dimension in truth discernment find conflicting results. Using data from Germany and the UK, Arin et al. (2023) find that women are less successful in detecting fake news than men. Similarly, Angelucci and Prat (2024) finds that women perform worse than men in detecting fake political news stories. In contrast, using Spanish data, Almenar et al. (2021) find no gender differences in the ability to detect fake news.[23] While Sindermann

[21]Our finding on the age dimension aligns with Pennycook and Rand (2019), Arin et al. (2023) and Angelucci and Prat (2024). In contrast to Arin et al. (2023) and Angelucci and Prat (2024), we find that education and household income does not predict US Americans' ability to detect fake news —if we simultaneously control for exogenous socioeconomic characteristics and the CRT and SCT scores.

[22]To control for media consumption, we build an *news media consumption* indicator. We asked subjects to rate the intensity of their media consumption for different media sources: 1. Television and/or radio; 2. Online social networks and/or messaging apps; 3. Online newspapers and news magazines; 4. Printed newspapers and news magazines. Subjects provide an answer on a scale from 1 to 5, where 1 corresponds to *"I am not familiar with this news source"* and 5 with *"It is a major source of news for me"*. Individual news media consumption is equal to the sum of the scale points across the four types of media. Therefore, *media consumption* $\in [4, 20]$. A higher score indicates higher frequency and/or broader use of sources (i.e., a more heterogeneous set).

[23]Concerning the consumption of fake news, Almenar et al. (2021) find gender differences in the topics of false information received. A higher proportion of men receive false news on political issues, whereas

et al. (2021) find that women perform better in detecting fake news than men. However, these gender differences are not statistically significant.

Differences in the selection of news topics might partially drive these opposing conclusions. The average woman's and the average man's preferences, interests, and, hence, knowledge differ on specific topics—especially politics. To avoid the news items tested being biased in a particular (gender) direction, we covered a wide range of news topics in our study. This allows us to investigate gender differences in detecting fake news. In this paper, we consistently find a significant gender difference; women perform better than men (Table 1, Columns 1–5).[24]

We investigate a possible explanation for this result—the role of gender differences in "skills" relevant for assessing the accuracy of news. In our setup, we focus exclusively on citizens' ability to discern the accuracy of the information based on the news content, not its visual aspects (e.g., fabricated pictures or videos). For this reason, all of our news items share the same appearance, with only text elements and no visual components. Hence, reading skills[25] and the ability to understand the meaning derived from a string of words (text comprehension) seems essential. We do not have a measure for the first skill, but we have a measure for the second skill, the SCT test. As Table 1 shows, the higher the SCT score, the better the ability to discern the truth. We find statistically significant gender differences in the SCT score, women perform better than men (Figure 5, Panel b).[26]

We summarize our findings on socioeconomic and behavioral determinants for the US population's susceptibility to fall for fake new as follows.

Result 2 (Heterogeneity in the Ability to detect Fake News.)**.** *The ability to distinguish between inaccurate and accurate news varies significantly along socioeconomic dimensions (gender, age), cognitive ability (CRT), text comprehension skills (SCT), political preferences, and type of media consumption. Women, older (+54), those who are more literate in text comprehension, those with*

women tend to receive fake news about celebrities more frequently. Wasike (2023) finds that women are more likely to comment on posts with misinformation than men.

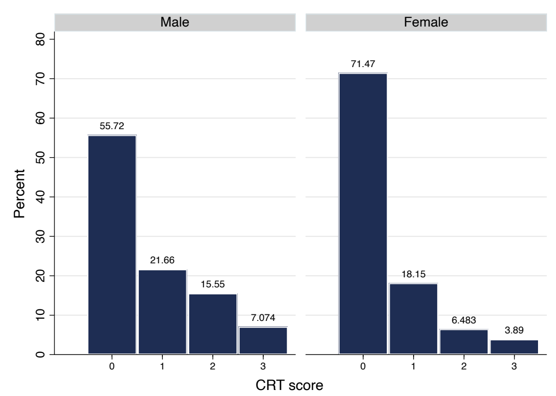
[24] Appendix Figure A2 shows the distribution of the performance score by gender, the gender differences in the mean ($t = -2.3054$, $p = 0.0212$), median (MWU-test, $z = -2.742$, $p = 0.0061$), and the distribution (Kolmogorov-Smirnov test) are statistically significant.

[25] There is abundant evidence that girls have better reading skills than boys. According to PISA reading scores, girls perform better than boys in all OECD countries (OECD, 2023). Halpern et al. (2007) find large gender differences in reading literacy. Fourth-grade girls performed significantly better than fourth-grade boys in the 33 countries. The advantage for women in reading literacy is replicated and is comparable in size to 15-year-olds in each of the 25 countries that participated in the Program for International Student Assessment.

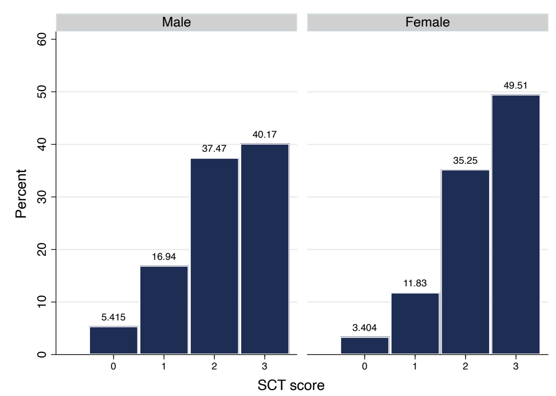
[26] Figure 5, Panel a, shows that the CRT score is higher for men than for women. The CRT test was introduced by Frederick (2005), who found that men significantly outperform women. Several subsequent studies have confirmed this result, using participants from different age groups, educational levels, and countries (using both the original CRT and modified versions; see Campitelli and Gerrans (2014); Cueva et al. (2016); Pennycook et al. (2016); Sinayev and Peters (2015); Ring et al. (2016); Albaity et al. (2014); Toplak et al. (2014); Primi et al. (2016)).

higher cognitive ability, those who trust fact-checking, those with democratic political preferences and those who do not use social media as the main source of news consumption perform significantly better in distinguishing between fake and accurate news.

Figure 5: Distribution of CRT and SCT scores (by gender)



(a) CRT Score



(b) SCT Score

Notes: Score Distribution in the Cognitive Reflection Test (CRT) and the Sentence Comprehension Test (SCT) by gender. The CRT test measures the extent to which individual cognition is based on reflective versus intuitive processes. Sentence comprehension refers to the ability to understand the meaning derived from a string of words, guided by a set of linguistic structures and constraints (e.g., syntax, semantics). The gender difference is highly significant for both test scores (CRT: MWU test, $z = 8.628$, $p = 0.0000$; SCT: MWU test, $z = -5.275$, $p = 0.0000$).

Economic Choices: Before the accuracy-detection task, we measure subjects' economic choices pre-experiment. Subjects must allocate their budget between consumption goods, health insurance, and misinformation insurance (i.e., a service like a fact checker; see Section 3.2). We find that subjects allocate 36.06% of their hypothetical 1000$ to purchasing consumption goods, 43.77% to health insurance and 20.17% to insurance against the risk of being harmed by made-up news and information. Hence, the willingness to pay for the misinformation insurance is a non-negligible component. Importantly, subjects who report that they have already been harmed by fake news and information intended to mislead the public (roughly 40% of our sample) allocate a significantly larger share of their budget to misinformation insurance (+ 4.39 p.p.) – potentially suggesting greater awareness of the risks and harms associated with the diffusion of misinformation. A greater prior allocation to misinformation insurance correlates positively with considering misinformation at least as a *"moderately big problem"* for the country. Spending on misinformation insurance negatively correlates with trust in others, educational attainment, and CRT and SCT test scores.

5 Experimental Results

The between-subject design of the experiment was carefully thought to be engaging and simple (as described by Stantcheva, 2023). Figure 6 shows the feedback of the participants given at the end of the survey.[27] We conclude that the survey experiment successfully engaged the participants. Appendix Table A5 shows that the sample is balanced between control and treatment groups (e.g., across socioeconomic dimensions such as gender, age, race, income, education and employment status.

This section explores the potential impact of a simple policy intervention—informing participants about their actual susceptibility to fall for fake news, on their belief updating, and their willingness to pay for misinformation insurance. In particular, we test two hypotheses described in Section 3.

Figure 6: Participants' Engagement (word cloud)

Notes: At the end of the survey, participants receive the following final question: *"Do you have any comments about our survey?"*; with an open answer text field. This Figure shows a word cloud with all the important words used in sizes related to the frequency with which they are used.

5.1 Treatment Effect on Beliefs

This section investigates whether and to what extent subjects' beliefs respond to information about their own ability to detect fake news. Treated subjects receive information on their fake news detection ability score after the first and before the second part of the task (Figure 3).

[27]For visual clarity, Appendix Figure A3 shows an enlarged version of Figure 6.

5.1.1 Accuracy of beliefs

First, we analyze whether individuals' beliefs about their fake news detection ability become more accurate as a result of the score revelation information treatment. To do so, we define two measures of the accuracy of beliefs for individuals i, POE_i, and MTE_i. Both measures compute the difference between subjects' quantitative beliefs about their ability (i.e., their beliefs about how many correct answers they have given) and their actual performance score, with MTE measuring this difference before and POE after the information provision experiment. To investigate whether the information provision causally affects the precision of individuals' beliefs about their own ability to detect fake news, we estimate the following specification:

$$POE_i = \beta_0 + \beta_1 Treat_i + \beta_2 MTE_i + \beta_3' \boldsymbol{X_i} + \epsilon_i, \tag{5.1}$$

where POE_i denotes the absolute value difference between individual i's posterior belief and his/her actual score in the second part of the task; $POE_i = |belief_i^{postior} - score_i^{part2}|$. $Treat_i$ is the variable of interest, the treatment dummy that equals one if the individual i received the score revelation treatment and zero otherwise. X_i denotes the vector of individual-specific controls.[28] And MTE_i denotes the absolute value difference between individual i's mid-task belief and his/her actual score in the first part of the task; $MTE_i = |belief_i^{midtask} - score_i^{part1}|$. The error term is denoted by ϵ_i.

Table 2 reports the estimation result of specification (5.1) in Column 3. The coefficient of the treatment dummy is negative and highly statistically significant. The random assignment of subjects to the treatment and control groups allows us to conclude that providing participants with information about their ability to detect fake news causally improves the accuracy of their beliefs about their ability. In other words, the difference between the posterior belief of the individuals (i.e., the perceived posterior ability) and the actual ability is significantly smaller for the treated subjects than for those in the control group. Hence, the average treated participant trusts the information provided and adjusts his/her ability beliefs accordingly.

To test the robustness of this result, we show the estimation results with varying controls in Column 1–2. We consider two additional specifications. In particular, the information provided on individuals' performance during the first part of the task could affect individuals' effort and performance in the second part—and, in turn, contribute to the observed treatment effect. Therefore, we include in Column 4 two additional controls, Δscore and Δeffort. The proxy for subjects' potential change in performance (Δscore) simply measures the difference in performance (score)

[28] The vector X_i includes gender, age, marital status, household income, educational attainment, employment status, categories for ethnicity (race), location of residence, a measure for cognitive ability—the Cognitive Reflection Test (CRT) score, and a measure for text comprehension ability—the Sentence Comprehension Test (SCT) score.

in the second versus first part of the task. To account for a potential change in individual effort (Δeffort), we use the difference in the time spent on the second versus the first part of the task. Column 4 shows that the estimated coefficient of interest β_1 remains unchanged when taking into account these factors.

Dependent variable: POE_i					
	(1)	(2)	(3)	(4)	(5)
Treatment (dummy)	-0.222***	-0.179***	-0.174***	-0.177***	-0.177***
	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)
MTE			0.303***	0.302***	0.344***
			(0.02)	(0.02)	(0.03)
Δ score				0.053***	0.065***
				(0.02)	(0.02)
Δ effort				0.000	-0.001
				(0.02)	(0.02)
Internet (dummy)				0.166	0.214
				(0.15)	(0.15)
MTE sign					
positive perception					-0.454***
					(0.11)
negative perception					-0.279***
					(0.11)
Mid-task belief					-0.036
					(0.02)
Constant	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes
Observations	2413	2224	2224	2224	2224
R^2	0.01	0.02	0.10	0.11	0.12

Notes. OLS estimates from Equation 5.1. Robust standard errors are reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. POE_i denotes the absolute value difference between the posterior belief of the individual i' and his/her actual score in the second part of the task. MTE measures the difference between subjects' quantitative mid-task beliefs about their ability (i.e., subjects' beliefs about how many correct answers they have given) and their actual performance score for the first part of the task. Δ score denotes the score difference between the second and first part of the task. Δ effort denotes the change in effort measured as the difference between the time spent on the second part and the time spent on the first part of the task. The internet dummy equals one if the subject answered *yes* to the following survey question: "Did you search the internet (via Google or otherwise) for any of the news headlines? Please be honest! You will get your compensation regardless of your response.", and zero otherwise (columns 2 and 3). *Mid-Task Belief* denotes the level of subjects' mid-task beliefs (column 3). *MTE sign* denotes a categorical variable that classifies the difference of individuals' mid-task beliefs minus their actual score of part 1 into three categories (correct, positive, negative). Controls include gender, age, marital status, household income, educational attainment, employment status, categories for ethnicity (race), location of residence, a measure for cognitive ability—the Cognitive Reflection Test (CRT) score, and a measure for text comprehension ability—the Sentence Comprehension Test (SCT) score.

Table 2: Accuracy of Beliefs (about fake news detection ability)

In the specification presented in column 5, we include additional variables. Variables that capture differences in individuals' initial (pre-treatment) beliefs about their ability. We include the level of their mid-task belief about their ability (i.e., subjects' pre-treatment beliefs about how many correct answers they have given). We also control for a variable identifying whether subjects have had a positive, negative, or correct pre-treatment perception of their ability ($MTEsign_i$). We define positive (negative) self-perception by comparing mid-task beliefs and the actual performance score during the first part of the news-accuracy detection task. A participant has a positive (negative) self-perception about his(her) ability if (s)he believes that (s)he has given more (less) correct answers than (s)he actually did. The results are robust to the inclusion of these additional variables. Column 5 shows that the coefficient of our treatment dummy remains unchanged in both value and significance levels. Therefore, we conclude that Hypothesis H1 is verified. We summarize our finding as follows:

Result 3 (Information about personal susceptibility to fake news causally affects belief accuracy (H1 holds))**.** *Treated subjects significantly adjust their ability beliefs according to the information they receive. The information (signal) induces a belief adjustment such that treated subjects hold more accurate beliefs about their ability to detect fake news, compared to the control group.*

5.1.2 Belief Updating

The previous section showed that providing participants with information about their true ability significantly narrows the gap between perceived and actual ability, even when accounting for changes in subjects' performance and effort between the two task segments. This result strongly indicates that the observed belief-ability gap reduction stems primarily from an exogenous shift in beliefs, induced by the information treatment. Therefore, in the following, our aim is to explore the dynamics of beliefs updating in more detail.

Magnitude and direction. We aim to explore whether the magnitude of the treatment matters for the magnitude of belief updating. In other words, if an individual learns that he was performing only half as well as believed, will this individual update his beliefs much more than someone who learns that his actual ability was only slightly below his believed ability? Second, does it depend on whether you receive a positive or negative signal (i.e., information) about your ability?

To answer these questions, we first define the variable $updating_i$ to quantify the magnitude of individual i's belief updating about his ability to detect fake news. The variable $updating_i$ denotes the quantitative difference between posterior and mid-task beliefs of how many correct answers have been given; $updating_i = belief_i^{posterior} - belief_i^{midtask}$ (Figure 3). Next, we define the variable $Treat_i^{Sign}$ that captures whether individual i received in the score revelation treatment a positive or negative signal (i.e., information) about his ability.[29]

[29] $Treat_i^{Sign}$ is a categorical variable and identifies whether individual i (i) is part of the control group, or whether individual i has received (ii) positive feedback, (iii) negative feedback, or (iv) feedback that his ability matches his beliefs.

Next, we investigate the effect of an interaction term between the augmented treatment identifier $Treat_i^{Sign}$ and MTE_i on $updating_i$. Recall, MTE_i denotes the pre-treatment belief-ability gap; i.e., the absolute value difference between the level of individual i's mid-task beliefs and his score in the first part of the task. MTE_i should be interpreted as a measure of the magnitude of the information feedback—for treated subjects. We estimate the following specification by ordinary least squares (OLS):

$$updating_i = \beta_0 + \beta_1 Treat_i^{Sign} \times MTE_i + \beta_3' \boldsymbol{X_i} + \beta_4' \boldsymbol{Z_i} + \epsilon_i, \tag{5.2}$$

where X_i denotes the vector of socioeconomic controls for individual i.[30] Z_i denotes a vector of individual i's characteristics revealed during the experiment, including: the change in subjects' performance (Δ score) and effort (Δ effort) between the second and the first part of the task, their level of mid-task beliefs, and finally $MTESign_i$, subjects' perception of their ability ($MTESign_i$).[31] The error term is denoted by ϵ_i.

For expositional clarity, Table 3 reports only the marginal effects for the interaction terms of interest and highlights two results.[32] First, the interaction terms are significant. Therefore, the magnitude of the treatment matters for the magnitude of belief updating. The adjustment in beliefs, resulting from the information treatment, is larger the less accurate individuals' prior ability beliefs. Second, we find that participants update their beliefs in the direction of the signal they receive: positive feedback recipients adjust beliefs upwards, while negative feedback recipients adjust downwards. However, the magnitude of the adjustment differs between the two groups—the adjustment observed among participants receiving negative feedback is nearly twice as large compared to those receiving positive feedback. In other words, participants who received the information that they overestimated their fake news detection ability (i.e., negative feedback) reacted stronger to the score revelation treatment than participants who received the information that underestimated their ability (i.e., positive feedback).

[30] As before, the vector X_i includes gender, age, marital status, household income, educational attainment, employment status, categories for ethnicity (race), location of residence, a measure for cognitive ability—the CRT score, and a measure for text comprehension ability—the SCT score.

[31] Recall that $MTEsign_i$ captures whether subject i has had a positive, negative, or correct pre-treatment perception of their ability. We define positive (negative) self-perception by comparing mid-task beliefs and the actual performance score during the first part of the news-accuracy detection task. A participant has a positive (negative) self-perception about his(her) ability if (s)he believes that (s)he has given more (less) correct answers than (s)he actually did.

[32] Appendix Table A7 reports all estimated OLS coefficients of specification (5.2).

Dependent variable: $updating_i = belief_i^{posterior} - belief_i^{midtask}$	
Positive feedback	0.135**
	(0.057)
Negative feedback	-0.222***
	(0.062)
Controls	yes
Observations	2413

Notes. Marginal effects for specification (5.2). Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors are in parentheses. Corresponding regression: $updating_i = \beta_0 + \beta_1 Treat_i^{Sign} \times MTE_i + \beta_3' \boldsymbol{X_i} + \beta_4' \boldsymbol{Z_i} + \epsilon_i$. The table shows the coefficients of interest, β_1. Socio-demographic controls X_i include gender, age, marital status, household income, educational attainment, employment status, categories for ethnicity (race), location of residence, a measure for cognitive ability—the Cognitive Reflection Test (CRT) score, and a measure for text comprehension ability—the Sentence Comprehension Test (SCT) score. Z_i denotes a vector of individual i's characteristics revealed during the experiment, including the change in subjects' performance (Δ score) and effort (Δ effort) between the second and first part of the task, their level of mid-task beliefs, and finally $MTESign_i$, subjects' "perception" of their ability ($MTESign_i$).

Table 3: Belief updating and feedback (information) sign

This is ideal from a policy perspective; those participants who were the most unaware of their personal susceptibility to fall for fake news reacted the most.[33] Appendix Table B1 shows that this result is robust to using qualitative beliefs instead of quantitative beliefs.

Result 4 (Magnitude of belief updating largest for negative feedback)**.** *The size of belief updating is stronger among subjects receiving negative feedback i.e., individuals who had an overoptimistic perception of their ability before receiving the signal about their actual ability to detect fake news.*

5.2 Treatment Effect on Economic Choices

Our findings have established that providing individuals with feedback on their ability to discern the veracity of news headlines effectively shifts their beliefs, leading to a more accurate assessment of their skills (Result 3). Notably, the magnitude of belief updating is greater among participants receiving negative feedback (Result 4). This finding shows that those who initially had an overinflated perception of their ability to detect fake news adjusted their ability beliefs more substantially. Finally, this Section tests Hypothesis H2 by investigating whether this enhanced self-awareness about their personal susceptibility to fall for fake news significantly changes subjects' willingness to hedge against the risk of misinformation. Specifically, Hypothesis 2 tests whether participants who received a signal (i.e., feedback that they hold inaccurate beliefs about their ability to detect

[33] We investigate whether these respondents (those who received negative feedback) update their beliefs consistent with Bayesian updating. In this case, we should observe stronger updating among subjects with lower confidence in mid-task beliefs. To explore this, we regressed $updating_i$ on MTE_i interacted with a dummy equal to one if subjects are at least "sure" about their mid-task beliefs. We find evidence for Bayesian learning: the magnitude of downward updating is larger among participants who are less confident in their mid-task beliefs than among those with greater confidence in their beliefs. Results are available on request.

fake news) and trusted the signal (hence, updated their ability beliefs) will significantly change their willingness to pay for misinformation insurance.

Belief updating and misinformation insurance. Before and after the information treatment, we measure subjects' economic choices. Subjects must allocate their budget between consumption goods, health insurance, and misinformation insurance (i.e., a service like a fact checker), see Section 3.2.[34] Let us denote the change in individual i's willingness to pay for misinformation insurance by Δwtp_i i.e., the difference between the posterior budget allocation decision and the prior ($\Delta wtp_i = post_i^{alloc} - prior_i^{alloc}$).

To test Hypothesis 2, we estimate the following specification:

$$\Delta wtp_i = \beta_0 + \beta_1 Treat_i \times updating_i + \beta_2' \boldsymbol{X_i} + \beta_3' \tilde{\boldsymbol{Z}}_{\boldsymbol{i}} + \epsilon_i, \tag{5.3}$$

where $Treat_i$ denotes the treatment dummy that equals one if the individual i received the score revelation treatment and zero otherwise. X_i denotes the vector of socioeconomic controls for individual i.[35] $updating_i$ denotes the quantitative difference between individual i's posterior and mid-task beliefs (i.e., beliefs about their fake news detection ability; $updating_i = belief_i^{posterior} - belief_i^{midtask}$). $\tilde{Z}_i$ denotes a vector of individual i's characteristics revealed during the experiment, including the change in subjects' performance (Δ score) and effort (Δ effort) between the second and first part of the task. To account for initial individual differences that could be relevant for this particular decision on how to allocate the budget, $\tilde{Z}_i$ also includes a proxy for individual i's risk aversion and health status.[36] Furthermore, this vector includes the initial percentage of individual i of the \$1000 budget allocated to the misinformation insurance (pre-experiment). The error term is denoted by ϵ_i.

Table 4 reports the results and highlights the average marginal effects of the ability belief updating on the change in spending Δwtp_i for the control group and the treatment group separately. The update (i.e., change) in the beliefs about the ability to detect fake news results in a change in economic choices for the treatment group. For the control group that does not receive any information, changes in fake news detection ability beliefs do not lead to statistically significant

[34] Pre-experiment, subjects allocate 36.06% of their hypothetical \$1000 to purchasing consumption goods, 43.77% to the health insurance, and 20.17% to the insurance against the risk of being harmed by made-up news and information. Hence, the willingness to pay for the misinformation insurance is a non-negligible component already pre-experiment.

[35] As before, the vector X_i includes gender, age, marital status, household income, educational attainment, employment status, categories for ethnicity (race), location of residence, a measure for cognitive ability—the CRT score, and a measure for text comprehension ability—the SCT score.

[36] To elicit respondents' risk aversion, we follow Dohmen et al. (2011) by asking about their general willingness to take risks on a scale from 0 to 10, where 0 indicates no willingness to take risks and 10 indicates intense risk-taking. The proxy for the health status is a dummy variable, denoted by hd_i, that equals one if individual i reports being in a *"good"* or *"very good"* health status.

changes in the willingness to pay for misinformation insurance.

Dependent variable: change in budget allocated to misinformation insurance Δwtp_i	
Control	-3.850
	(5.846)
Treatment	-6.633**
	(3.340)
Observations	2413

Notes. Marginal effects for specification (5.3). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors are in parentheses. Corresponding regression: $\Delta wtp_i = \beta_0 + \beta_1 Treat_i \times updating_i + \beta_2' \boldsymbol{X_i} + \beta_3' \tilde{\boldsymbol{Z}}_{\boldsymbol{i}} + \epsilon_i$. The table shows the marginal effect of the coefficients of interest, β_1. The socio-demographic controls X_i include gender, age, marital status, household income, educational attainment, employment status, categories of ethnicity (race), location of residence, a measure of cognitive ability – the Cognitive Reflection Test (CRT) score, and a measure of text comprehension ability – the Sentence Comprehension Test (SCT) score. $\tilde{Z}_i$ denotes a vector of individual i's characteristics revealed during the experiment, including: the change in subjects' performance (Δ score) and effort (Δ effort) between the second and the first part of the task, a proxy for individual i's risk aversion and health status, and individual i's initial percentage of the \$1000 budget allocated to the misinformation insurance (pre-experiment).

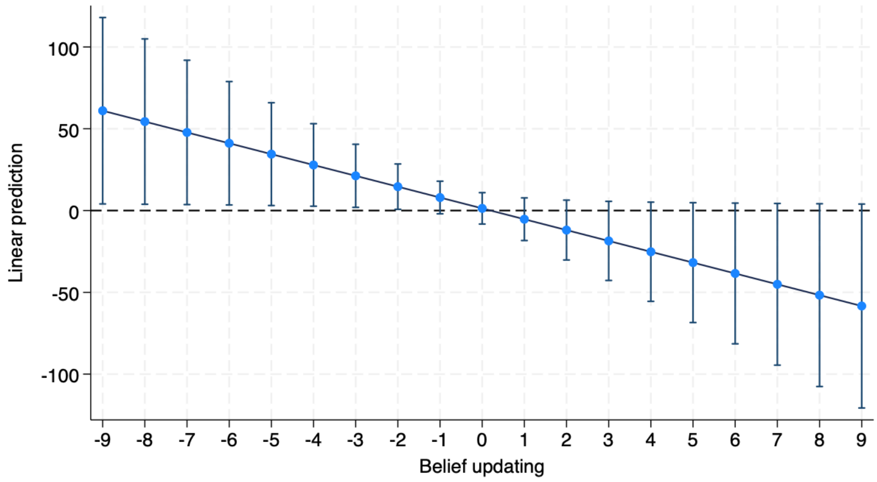
Table 4: Belief updating and misinformation insurance: marginal effects

Note that the sign and magnitude of this average treatment effect reported in Table 4 are not very informative. The previous Section showed that the direction (Result 3) and magnitude (Result 4) of the information treatment effect on beliefs depend on whether the subject received positive or negative feedback concerning his ability to detect fake news. Treated subjects significantly adjust their ability beliefs according to the information treatment they receive. Those who overestimated their fake news detection ability (i.e., received negative feedback) adjusted their beliefs downward and vice versa. In terms of magnitude, we found that the effect of treatment on ability beliefs is more substantial for those who receive negative feedback. Treated subjects who learned they overestimated their ability to detect fake news adjusted their beliefs more strongly. Therefore, we investigate the treatment effect on the willingness to pay for misinformation insurance for each level of ability belief updating separately.

Figure 7 illustrates our key result and highlights substantial heterogeneity in the treatment effect. Figure 7 visualizes the empirical evidence for Hypothesis H2 by showing that the (average) estimated treatment effect on the willingness to pay is positive and significant only for subjects who experience a negative belief update.[37] Notably, Figure 7 reveals a linear relationship. The stronger this downward adjustment in ability beliefs, the larger the increase in the amount allocated to misinformation insurance.

[37] Considering the distribution of ability belief updating among these subjects, the weighted average increase in their spending on misinformation insurance amounts to \$14.8, corresponding to a +7% increase compared to the initial amount (pre-experiment).

Figure 7: Treatment effect on Economic Choices (by level of ability belief $updating_i$)



Notes: Dots and vertical bars represent the linear regression coefficients (marginal effects) and their confidence intervals, respectively. The depending variable is the change in individual i's willingness to pay for misinformation insurance, i.e., the difference between the posterior budget allocation decision and the prior ($\Delta wtp_i = post_i^{alloc} - prior_i^{alloc}$). The independent variable belief updating, denoted by $updating_i$, is the quantitative difference between individual i's elicited posterior (after the information provision) and mid-task (before the information provision) fake news detection ability beliefs—beliefs of how many correct answers have been given in incentivized news-accuracy detection task; $updating_i = belief_i^{posterior} - belief_i^{midtask}$.

In summary, the simple intervention of providing citizens with information about their personal susceptibility to fall for fake news, i.e., more precisely, information about their actual ability to distinguish between accurate and inaccurate news, leads them to revise their ability beliefs and their economic choices—their willingness to pay to protect themselves from the harms of misinformation.

Since the estimated average treatment effect on economic choices stems exclusively from subjects who exhibit downward belief adjustments, those who significantly increase their spending to hedge against misinformation are individuals who initially held an overly optimistic perception of their fake news detection ability. Upon receiving negative feedback, these individuals significantly review their beliefs, recognizing their limitations in distinguishing accurate news from fake news. Therefore, our final result can be summarized as follows:

Result 5 (Awareness of personal susceptibility to fall for fake news leads to higher willingness to pay for services such as fact checking (H2 holds))**.** *The treated subjects who receive a negative signal about their fake news detection ability, update their beliefs significantly. This downward adjustment in ability beliefs increases subjects' willingness to pay for misinformation insurance (i.e., a service such as a fact checker) — compared to the control group.*

6 Conclusion

The proliferation of fake news has become a pressing concern, with potentially dire consequences for individuals and society. This paper investigates the critical question of whether citizens are aware of their personal susceptibility to fall for fake news. In addition, would citizens use information about their own fake news detection ability to update their beliefs about their susceptibility to fall for fake news? Does this belief updating affect their economic choices to protect themselves from misinformation? This paper fills these gaps in the literature.

To this end, we conducted a survey experiment using a representative sample of the US population. Instead of focusing on political fake news (as in Angelucci and Prat, 2024 and Guriev et al., 2023 among others), statements of a single politician (as in Barrera et al., 2020) or a specific topic (as in Lutzke et al., 2019 on fake news about climate change or Arechar et al. (2023) on the Covid-19 pandemic), we cover a heterogeneous spectrum of news topics and news outlets. Our approach focuses exclusively on the ability of citizens to discern the accuracy of information based on the news content, not its visual aspects (e.g., fabricated pictures, grammatical errors).[38]

We find that most survey participants consider fake news a problem for the country and only 37.61% believe that the average American can identify news or information that misrepresents reality or is even false. However, the vast majority have tremendous confidence in their own ability to evaluate news accuracy; 82.64% report having a "good" or "very good" ability to identify news or information that misrepresents reality or is even false.

Academics and policymakers discuss a variety of interventions to address the problem of fake news. One approach focuses on regulation, urging social media platforms to moderate and verify news content. However, as AI develops rapidly, existing detection tools must be more accurate to act as a reliable and definitive safeguard.[39] Other potential policies to address fake news include improving educational outcomes or providing specific training in media literacy and competency. These interventions would likely improve citizens' fake news detection ability, but require serious financial investments and planning horizons. In contrast, this paper provides evidence for the effectiveness of a simple and low-cost intervention, at least for those who trust fact-checking companies in the first place.

[38] The news items are heterogeneous along two essential dimensions: the publisher and the topic. Half of our news items appeared on mainstream news sources (e.g., FOX News, the Wall Street Journal, the New York Times) and the other half on alternative news sources (e.g., Naturalnews.com, Breitbart, Raw Story). In addition, we select news items covering a wide range of topics, such as crime, science, politics, the Covid-19 pandemic, climate change, and health.

[39] One drawback of this approach is that it might create political tensions. According to the Financial Times, the efforts of US-based tech groups to invest in fact-checking and tackling misinformation have also become politicized, as right-wing US politicians accuse them of colluding with the government and academics to censor conservative views.

We implement a policy intervention that educates citizens about their personal susceptibility to fall for fake news, and we demonstrate that those who receive a negative signal about their fake news detection ability, significantly adjust their beliefs accordingly. This downward shift in ability beliefs bolsters subjects' willingness to pay for misinformation insurance (i.e., a service such as a fact checker).

Our findings highlight the importance of understanding people's beliefs about their ability to assess the accuracy of information and the potential benefits of interventions that can improve awareness of these limitations. By promoting more accurate beliefs, we can empower individuals to make better decisions and mitigate the detrimental effects of misinformation (at least among those subjects who trust non-partisan fact-checkers).

Our experimental findings provide compelling evidence for the effectiveness of straightforward policy interventions in enhancing awareness of personal susceptibility to fall for fake news, and by that diminishing the proliferation of fake news. As we continue to navigate the complexities of the digital age, such evidence-based approaches will prove increasingly essential in safeguarding our society from the insidious influence of misinformation.

Bibliography

Acemoglu, D., Ozdaglar, A., and Siderius, J. (2023). A model of online misinformation. *The Review of Economic Studies*, forthcoming.

Albaity, M., Rahman, M., and Shahidul, I. (2014). Cognitive reflection test and behavioral biases in malaysia. *Judgment Decision Making*, 9(2):149–151.

Alicke, M. and Govorun, O. (2005). The better-than-average effect. In Alicke, M. D., Dunning, D. A., and Krueger, J. I., editors, *The Self in Social Judgment*, pages 85–106. Psychology Press.

Allcott, H. and Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–236.

Almenar, E., Aran-Ramspott, S., Suau, J., and Masip, P. (2021). Gender differences in tackling fake news: Different degrees of concern, but same problems. *Media and Communication*, 9(1).

Anderson, C., Brion, S., Moore, D. A., and Kennedy, J. A. (2012). A status-enhancement account of overconfidence. *Journal of Personality and Social Psychology*, 103(4):718–735.

Angelucci, C. and Prat, A. (2024). Is Journalistic Truth Dead? Measuring How Informed Voters Are about Political News. *American Economic Review*.

Apuke, O. D. and Omar, B. (2021). Social media affordances and information abundance: Enabling fake news sharing during the covid-19 health crisis. *Health Informatics Journal*, 27(3):14604582211021470. PMID: 34250868.

Arechar, A., Allen, J., Berinsky, A., and etal. (2023). Understanding and combatting misinformation across 16 countries on six continents. *Nature Human Behavior*, 7:1502–1513.

Arin, K. P., Mazrekaj, D., and Thum, M. (2023). Ability of detecting and willingness to share fake news. *Nature: Scientific Reports*, 13(7298).

Bago, B., Rand, D. G., and Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General*, 149(8):1608–1613.

Barrera, O., Guriev, S., Henry, E., and Zhuravskaya, E. (2020). Facts, alternative facts, and fact checking in times of post-truth politics. *Journal of Public Economics*, 182:1–18.

Benabou, R. (2013). Groupthink: Collective Delusions in Organizations and Markets. *The Review of Economic Studies*, 80(2):429–462.

Benabou, R. and Tirole, J. (2011). Identity, Morals, and Taboos: Beliefs as Assets *. *The Quarterly Journal of Economics*, 126(2):805–855.

Bermes, A. (2021). Information overload and fake news sharing: A transactional stress perspective exploring the mitigating role of consumers' resilience during covid-19. *Journal of Retailing and Consumer Services*, 61.

Bottan, N. L. and Perez-Truglia, R. (2022). Choosing your pond: Location choices and relative income. *The Review of Economics and Statistics*, 104(5):1010–1027.

Campitelli, G. and Gerrans, R. (2014). Does the cognitive reflection test measure cognitive reflection? a mathematical modeling approach. *Memory Cognition*, 42:434–447.

Charness, G., Gneezy, U., and Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior Organization*, 81(1):1–8.

Chiang, C.-F. and Knight, B. (2011). Media bias and influence: Evidence from newspaper endorsements. *The Review of Economic Studies*, 78(3):795–820.

Chopra, F., Haaland, I., and Roth, C. (2022). Do People Demand Fact-Checked News? Evidence from U.S. Democrats. *Journal of Public Economics*, 205.

Christelis, D., Georgarakos, D., Jappelli, T., Pistaferri, L., and Van Rooij, M. (2019). Asymmetric consumption effects of transitory income shocks. *Economic Journal*, 129(622):2322–2341.

Cueva, C., Iturbe-Ormaetxe, I., Mata-Pérez, E., Ponti, G., Sartarelli, M., Yu, H., and Zhukova, V. (2016). Cognitive (ir)reflection: New experimental evidence. *Journal of Behavioral and Experimental Economics*, 64:81–93.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.

Dunning, D., Heath, C., and Suls, J. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychol Sci Public Interest*, 5(3).

Enyedi, Z. and Kreko, P. (2018). Explaining eastern europe: Orbán's laboratory of illiberalism. *Journal of Democracy*, 29(3):39–51.

Eurobarometer (2018). Flash eurobarometer survey 464: Fake news and disinformation online.

Faia, E., Fuster, A., Pezone, V., and Zafar, B. (2022). Biases in information selection and processing: Survey evidence from the pandemic. *The Review of Economics and Statistics*, forthcoming.

Fraile, M. and Gomez, R. (2017a). Bridging the enduring gender gap in political interest in europe: The relevance of promoting gender equality. *European Journal of Political Research*, 56.

Fraile, M. and Gomez, R. (2017b). Why Does Alejandro Know More about Politics than Catalina? Explaining the Latin American Gender Gap in Political Knowledge. *British Journal of Political Science*, 47(1):91–112.

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, 19(4):25–42.

Fuster, A., Kaplan, G., and Zafar, B. (2021). What Would You Do with $500? Spending Responses to Gains, Losses, News, and Loans. *The Review of Economic Studies*, 88(4):1760–1795.

Gerber, A. S., Gimpel, J. G., Green, D. P., and Shaw, D. R. (2011). How large and long-lasting are the persuasive effects of televised campaign ads? results from a randomized field experiment. *American Political Science Review*, 105(1):135–150.

Gerber, A. S., Karlan, D., and Bergan, D. (2009). Does the media matter? a field experiment measuring the effect of newspapers on voting behavior and political opinions. *American Economic Journal: Applied Economics*, 1(2):35–52.

Gleick, J. (2011). *The Information: A History, A Theory, A Flood.* Pantheon Books.

Guriev, S., Henry, E., Marquis, T., and Zhuravskaya, E. (2023). Curtailing False News, Amplifying Truth. Working Papers halshs-04315924, HAL.

Haaland, I. K., Roth, C., and Wohlfart, J. (2023). Designing information provision experiments. *Journal of Economic Literature*, 61(1):3–40.

Halpern, D., Benbow, C., Geary, D., Gur, R., Hyde, J., and Gernsbacher, M. (2007). The science of sex differences in science and mathematics. *Psychological Science Public Interest*, 8(1):1–51.

Henry, E., Zhuravskaya, E., and Guriev, S. (2022). Checking and Sharing Alt-Facts. *American Economic Journal: Economic Policy.*

Jappelli, T. and Pistaferri, L. (2014). Fiscal policy and MPC heterogeneity. *American Economic Journal: Macroeconomics*, 6(4):107–136.

Kang, M. J., Rangel, A., Camus, M., and Camerer, C. F. (2011). Hypothetical and real choice differentially activate common valuation areas. *The Journal of Neuroscience*, 12:461–468.

Lawson, M., Anand, S., and Kakkar, H. (2023). ribalism and tribulations: The social costs of not sharing fake news. *Journal of Experimental Psychology: General*, 152(3):611–631.

Lutzke, L., Drummond, C., Slivic, P., and Arvai, J. (2019). Priming critical thinking: Simple interventions limit the influence of fake news about climate change on Facebook. *Global Environmental Change*, 58.

Lyons, B. A., Montgomery, J. M., Guess, A. M., Nyhan, B., and Reifler, J. (2021). Overconfidence in news judgments is associated with false news susceptibility. *Proceedings of the National Academy of Sciences*, 118(23).

Maertens, R., Götz, F., Golino, H., and et al. (2023). The Misinformation Susceptibility Test (MIST): A psychometrically validated measure of news veracity discernment. *Behavior Research Methods*.

Moore, D. A. and Healy, P. J. (2008). The Trouble With Overconfidence. *Psychological Review*, 115(2):502–517.

Nathan, B. C., Perez-Truglia, R., and Zentner, A. (2020). My taxes are too darn high: Why do households protest their taxes? Working Paper 27816, National Bureau of Economic Research.

OECD (2023). Reading performance (pisa) (indicator). doi: 10.1787/79913c69-en (accessed on 30 november 2023).

Osmundsen, M., Bor, A., VAHLSTRUP, P. B., BECHMANN, A., and PETERSEN, M. B. (2021). Partisan Polarization Is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter. *American Political Science Review*, 115(3):999–1015.

Parker, J. A. and Souleles, N. S. (2019). Reported Effects versus Revealed-Preference Estimates: Evidence from the Propensity to Spend Tax Rebates. *American Economic Review: Insights*, 1(3):273–290.

Parker, J. A., Souleles, N. S., Johnson, D. S., and McClelland, R. (2013). Consumer spending and the economic stimulus payments of 2008. *American Economic Review*, 103(6):2530–2553.

Pennycook, G., Cheyne, J. A., Koehler, D. J., and Fugelsang, J. A. (2016). Is the cognitive reflection test a measure of both reflection and intuition? *Behavior Research Methods*, 48(1):341–348.

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., and Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592(7855):590–595.

Pennycook, G. and Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188:39–50. The Cognitive Science of Political Thought.

Pennycook, G. and Rand, D. G. (2020a). Who falls for fake news? the roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*, 88(2):185–200.

Pennycook, G. and Rand, D. G. (2020b). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*, 88(2):185–200.

Pew (2016a). Pew research survey: Many americans believe fake news is sowing confusion.

Pew (2016b). Pew research survey: nformation overload.

Pew (2019). Pew research survey: Many americans say made-up news is a critical problem that needs to be fixed.

Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., and Hamilton, J. (2016). The development and testing of a new version of the cognitive reflection test applying item response theory (irt). *Journal of Behavioral Decision Making*, 29:453–469.

Quaranta, M. and Dotti Sani, G. M. (2018). Left Behind? Gender Gaps in Political Engagement Over the Life Course in Twenty-Seven European Countries. *Social Politics: International Studies in Gender, State Society*, 25(2):254–286.

Ring, P., Neyse, L., David-Barett, T., and Schmidt, U. (2016). Gender differences in performance predictions: Evidence from the cognitive reflection test. *Frontiers in Psychology*, 7.

Ross, R. M., Rand, D. G., and Pennycook, G. (2021). Beyond "fake news": Analytic thinking and the detection of false and hyperpartisan news headlines. *Judgment and Decision Making*, 16(2):484–504.

Roth, C. and Wohlfart, J. (2020). How do expectations about the Macroeconomy affect personal expectations and Behavior? *The Review of Economics and Statistics*, 102(4):731–748.

Sinayev, A. and Peters, E. (2015). Cognitive reflection vs. calculation in decision making. *Frontiers in Psychology*, 6.

Sindermann, C., Schmitt, H. S., Rozgonjuk, D., Elhai, J. D., and Montag, C. (2021). The evaluation of fake and true news: on the role of intelligence, personality, interpersonal trust, ideological attitudes, and news consumption. *Heliyon*, 7(3):e06503.

Stantcheva, S. (2023). How to run surveys: A guide to creating your own identifying variation and revealing the invisible. *Annual Review of Economics*, 15:205–234.

Statista (2020). Survey on the level of confidence in distinguishing between real news and fake news among adults in the united states as of december 2020.

Stiglitz, J. E. and Kosenko, A. (2024). The economics of information in a world of disinformation: A survey part 2: Direct communication. Working Paper 32050, National Bureau of Economic Research.

Toplak, M. E., West, R. F., and Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the cognitive reflection test. *Thinking Reasoning*, 20(2):147–168.

Tworzecki, H. (2019). Poland: A case of top-down polarization. *The ANNALS of the American Academy of Political and Social Science*, 681(1):97–119.

Vernice, M., Matta, M., Tironi, M., Caccia, M., Lombardi, E., Guasti, M. T., Sarti, D., and Lang, M. (2019). An Online Tool to Assess Sentence Comprehension in Teenagers at Risk for School Exclusion: Evidence From L2 Italian Students. *Frontiers in Psychology*, 10.

Wasike, B. (2023). You've been fact-checked! examining the effectiveness of social media fact-checking against the spread of misinformation. *Telematics and Informatics Reports*, 11:100090.

YouGov (2023). Survey on misinformation susceptibility among adults in the united states as of april 2023.

Zimmermann, F. (2020). The dynamics of motivated beliefs. *American Economic Review*, 110(2):337–363.

Zizzo, D. (2010). Experimenter demand effects in economic experiments. *Experminental Econonomics*, 13(1):75–98.

Fake News:

Susceptibility, Awareness and Solutions

Online Supplementary Material

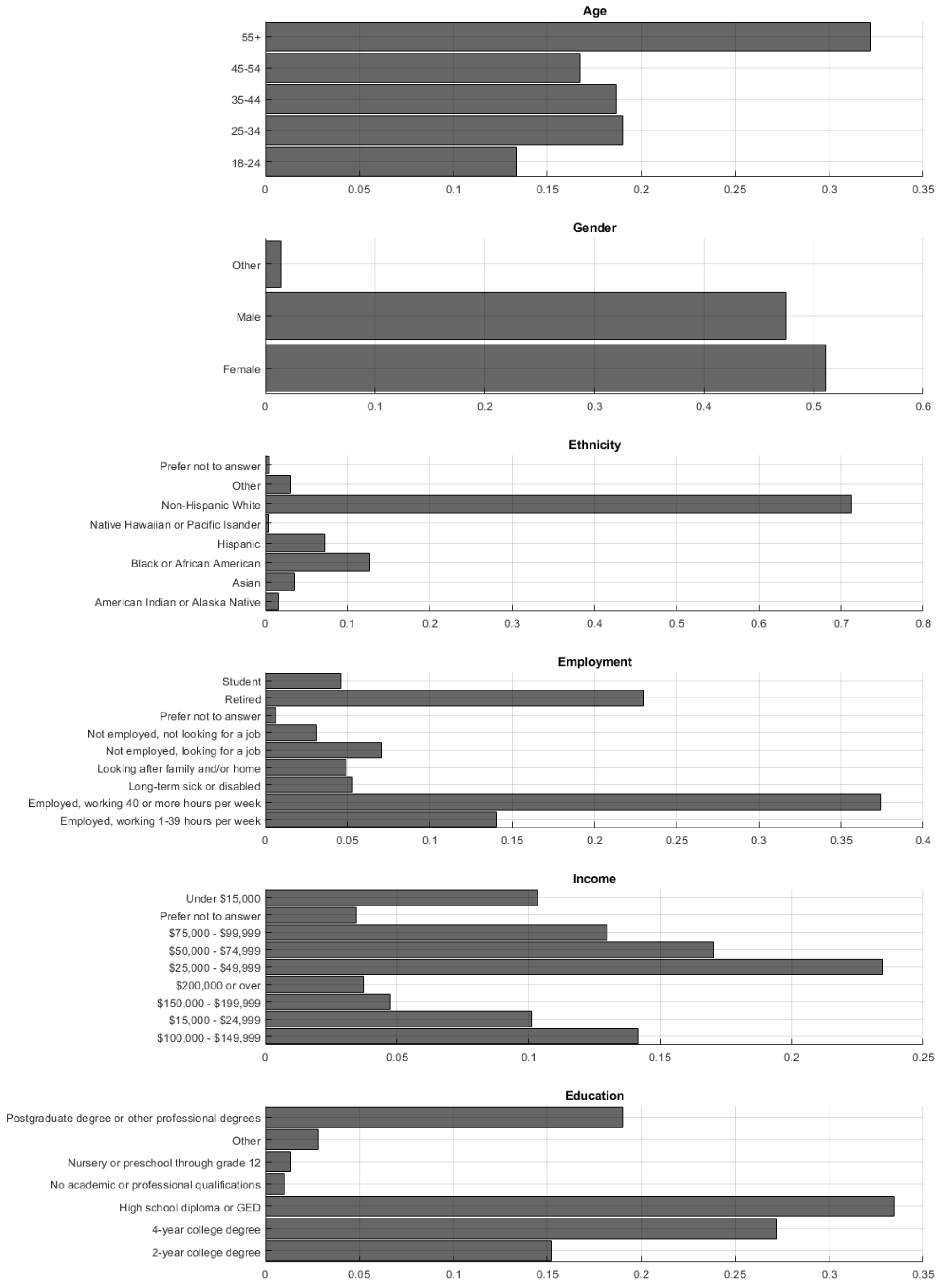
Section A provides descriptive statistics of the data, documents the representativeness of our sample, and reports the balance of sample test, and additional details for the estimations in the main text of the paper.

Figure A1 shows the sample's distribution of the key socioeconomic characteristics. Table A1 compares the key socioeconomic characteristics of our samples to nationally representative statistics from the U.S. Census Bureau. Our sample matches the nationally representative statistics when comparing the gender distribution between men and women, the age distribution, household size, marital status, the employment rate, the share of higher educational attainment, as well as the region of residence. In addition, table A1 shows the representativeness of our sample along the income dimension. We match the income distribution rather well except for the highest income category (last decile). The underrepresentation of top incomes (the missing rich) is often the case in household surveys and, hence, a well-known issue in the household finance literature. Tables A2–A4 report descriptive statistics of the variables used in the main text of this paper. Table A5 reports the balance of sample test.

Figure A2 shows descriptive results of participant's performance—by gender. And Table A6 reports the estimates for the baseline specification in (4.1). Figure A3 shows an enlarged version of Figure 6 in the main text of the paper. Table A7 reports the OLS estimates of specification (5.2). Section B reports additional results and robustness checks.

A Appendix: Descriptive Statistics & Results

Figure A1: Descriptive Statistics (Socioeconomic Characteristics)



	US Population	Survey
Socioeconomic		
Female	51%	51.1%
Median age	38.8	43
Never married	34.2%	31.9%
Average household size	2.5	2.7
Employment rate	58.6%	51.5%
Education: High school graduate or higher	88.6%	94.9%
Income		
\$0–\$14,999	9.9%	10.3%
\$15,000–\$24,999	7.5%	10.1%
\$25,000–\$49,999	19.1%	23.5%
\$50,000–\$74,999	16.8%	17%
\$75,000–\$99,999	12.8%	12.9%
\$100,000–\$149,999	16.3%	14.2%
\$150,000–\$199,999	7.9%	4.7%
\$200,000+	9.8%	3.7%
Geographic Area		
Northeast	17.22%	18.53%
Midwest	20.74%	22.55%
South	38.33%	39.63%
West	23.70%	19.29%
Ethnicity		
White	58.1%	71.2%
Black/African American	12.1%	12.7%

Notes. The table reports U.S. representative statistics from the Census Bureau (https://data.census.gov/profile/United_States?g=010XX00US)(column 1) alongside summary statistics from our survey (column 2). Socio-economic data for US population are from Census Bureau 2021 American Community survey 1 year estimates (https://data.census.gov/table/ACSDP1Y2021.DP02?g=010XX00US; https://data.census.gov/table/ACSDP1Y2021.DP05?g=010XX00US). Share of females in US population and survey is calculated for 18 years old and over. The median age in US population is determined over the total population, while in our survey is calculated over the population of 18 years old and over only (77.9% of the total population according to Census Bureau 2021 American Community survey 1 year estimates). Marital status in US population is calculated for 15 years old and over, while in our survey is calculated over the population of 18 years old and over only. Employment rate in US population is calculated for 16 years old and over while in our survey we include 18 years old and over. Education in US population and survey is calculated for 25 years old and over (https://data.census.gov/table/ACSST1Y2021.S1501?g=010XX00US). Income data for US population are from U.S. Census Bureau. "INCOME IN THE PAST 12 MONTHS (IN 2021 INFLATION-ADJUSTED DOLLARS)." American Community Survey, ACS 1-Year Estimates Subject Tables, Table S1901, 2021 (https://data.census.gov/table/ACSST1Y2021.S1901?g=010XX00US). Geographic area for US population are Annual Estimates of the Resident Population for the United States, Regions, States, District of Columbia, and Puerto Rico: April 1, 2020 to July 1, 2021 (NST-EST2021-POP). Ethnicity data for US population are from 2020 Decennial Census (https://data.census.gov/table/DECENNIALDHC2020.P9?g=010XX00US).

Table A1: Sample characteristics (Representativeness of the Sample)

	Behavioral Characteristics B				*News Consumption I*			
	trust		fake news					
	general	fact checker	problem		social media	television	online	printed
Strongly disagree	140	116	62	I am not familiar with this news source	54	15	68	53
Disagree	572	179	58	Is never a source of news for me	555	112	428	
Neither agree, nor disagree	619	625	329	Is rarely a source of news for me	406	289	476	596
Agree	940	1158	1026	Is a minor source of news for me	668	588	772	644
Strongly agree	142	335	938	Is a major source of news for me	730	1409	669	451
Total	2,413	2,413	2,413		2,413	2,413	2,413	2,413

The behavioral variables B include individual's trust in fact checking (5-point answer scale: strongly distrust to strongly trust), general trust in people (5-point answer scale: strongly distrust to strongly trust), perception of fake news being a problem for the country (4-point answer scale: very big problem to not a problem at all). Media consumption I variables include measures for the consumption frequency of four different news sources: social media, television, online newspapers, printed newspapers; for each (5-point answer scale, not familiar with this source of news to major source of news).

Table A2: Descriptive Statistics (Behavioral Characteristics and News Consumption I)

Variable	Obs	Mean	Std. dev.	Min	Max
Behavioral Characteristics B					
trust (general, people)	2,413	3.15	1.03	1	5
trust (fact checking)	2,413	3.59	0.98	1	5
Fake news, a problem for country	2,413	4.13	0.91	1	5
News Consumption I					
social media	2,413	3.61	1.20	1	5
television	2,413	4.35	0.91	1	5
online newspapers	2,413	3.64	1.14	1	5
printed newspapers	2,413	3.36	1.10	1	5

Notes. The behavioral variables B include individual's trust in fact checking (5-point answer scale: strongly distrust to strongly trust), general trust in people (5-point answer scale: strongly distrust to strongly trust), perception of fake news being a problem for the country (4-point answer scale: very big problem to not a problem at all). Media consumption I variables include measures for the consumption frequency of four different news sources: social media, television, online newspapers, printed newspapers; (5-point answer scale, not familiar with this source of news to major source of news).

Table A3: Descriptive Statistics (Behavioral Characteristics and News Consumption II)

Variable	Obs	Mean	Std. dev.	Min	Max
risk	2,413	5.45	2.55	0	10
CRT score	2,413	0.58	0.89	0	3
SCT score	2,413	2.22	0.85	0	3
health	2,413	3.95	0.78	1	5

Notes. To elicit respondents' *risk* aversion, we follow Dohmen et al. (2011) by asking about their general willingness to take risks on a scale from 0 to 10, where 0 indicates no willingness to take risks and 10 indicates intense risk-taking. To measure subjects' ability to comprehend sentences with varying levels of syntactic complexity, we use the *SCT* test developed by Vernice et al. (2019). For the Cognitive Reflection Test (*CRT*), we use the test developed by Frederick (2005). *Health* elicits respondents' perception of their household's health status using the survey question "How would you evaluate the overall health of the members of your household (including yourself)?" 5-point answer scale: Very good; Good; Fair; Bad; Very bad.

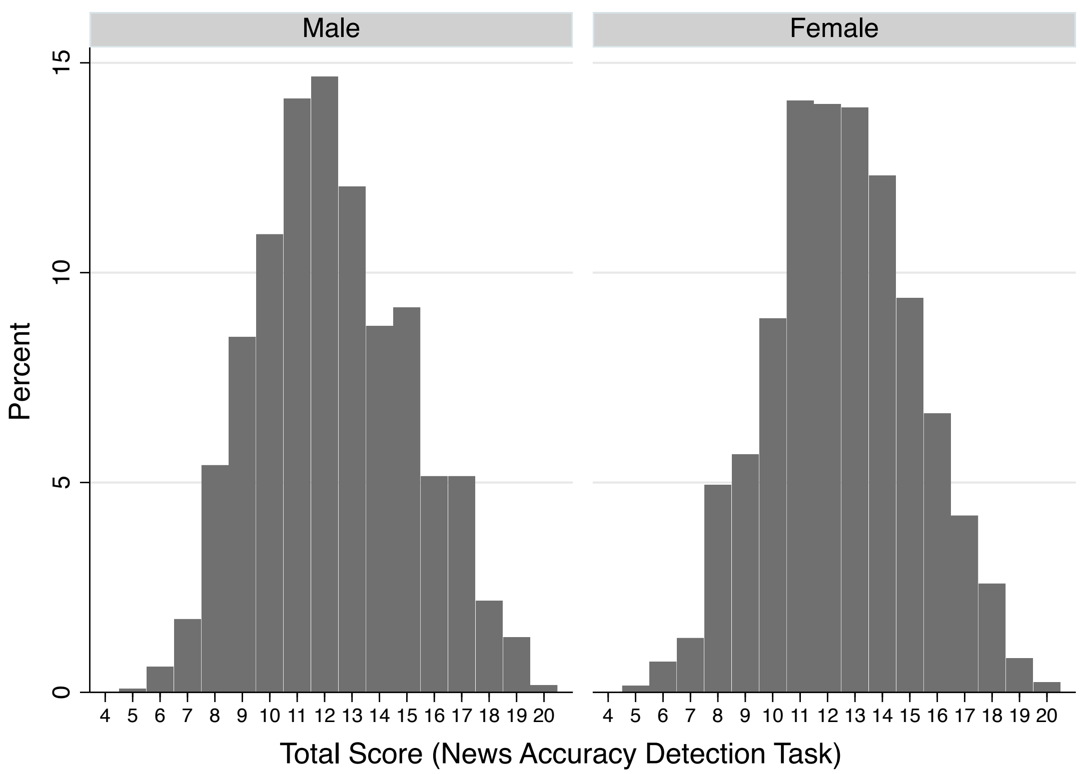
Table A4: Descriptive Statistics (CRT and SCT scores, risk and health)

	(1) Control	(2) Treatment	(3) P-value
Age	46.54	46.16	0.4083
Women	616	618	0.5029
Non-hispanic white	858	860	0.9355
Married or domestic partnership	630	617	0.5392
Income ($25,000 - $49,999)	283	283	0.7725
High-school diploma	392	415	0.4033
Employed, working 40 or more hours per week	449	454	0.9597
Observations	1206	1207	

Notes. Column 1 (2) reports the number of subjects in the control (treatment) group for the most frequent realization of each socio-demographic variable. For the variable Age, the average value is reported. Column 3 reports the p-value from the Wilcoxon rank-sum test.

Table A5: Balance of Sample Test

Figure A2: Distribution of the news-accuracy detection task score (by gender)



Dependent variable: Total Score					
	(1)	(2)	(3)	(4)	(5)
Female	0.277**	0.226*	0.307***	0.295***	0.212*
	(0.11)	(0.12)	(0.11)	(0.11)	(0.11)
middle aged	0.037	0.154	0.099	-0.014	0.203
	(0.13)	(0.14)	(0.13)	(0.13)	(0.14)
old	1.321***	1.169***	1.468***	1.086***	1.069***
	(0.14)	(0.16)	(0.14)	(0.15)	(0.17)
score sct	0.661***	0.590***	0.642***	0.623***	0.541***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
score crt	0.473***	0.478***	0.433***	0.447***	0.422***
	(0.07)	(0.07)	(0.06)	(0.07)	(0.06)
single		0.554***			0.550***
		(0.12)			(0.12)
currently employed		-0.409***			-0.456***
		(0.13)			(0.13)
middle education		0.029			-0.041
		(0.13)			(0.13)
high education		-0.068			-0.267
		(0.17)			(0.17)
income middle		0.204			0.157
		(0.13)			(0.13)
income high		0.041			0.069
		(0.18)			(0.18)
living in suburbs		0.401***			0.506***
		(0.13)			(0.13)
living in countryside		0.166			0.335**
		(0.17)			(0.17)
Behavioral factors B					
trust in fact checking			0.505***		0.535***
			(0.11)		(0.12)
fake news: a problem for country (perception)			0.189		0.212
			(0.13)		(0.13)
general trust (in people)			0.009		0.054
			(0.11)		(0.11)
political preference: Independent			-0.059		-0.247*
			(0.13)		(0.14)
political preference: Prefer not to answer			-0.530**		-0.695***
			(0.23)		(0.24)
political preference: Republican			-0.928***		-1.029***
			(0.13)		(0.14)
Media Consumption I					
Media Consumption Indicator				0.095**	0.072
				(0.05)	(0.05)
social media consumption				-0.858***	-0.713***
				(0.15)	(0.16)
television consumption				-0.109	-0.162
				(0.18)	(0.18)
online newspaper consumption				0.218	0.115
				(0.15)	(0.15)
printed newspaper consumption				-0.377***	-0.281*
				(0.14)	(0.15)
N	2324	2224	2324	2324	2224
R2	0.16	0.18	0.19	0.17	0.23

Notes: Columns 1-4 report OLS estimates. Robust standard errors (Eicker-Huber-White) are reported in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Socio-demographic exogenous controls include gender, age (young, middle aged, old), and categories for ethnicity (Non-Hispanic White, Hispanic, Black or African American, American Indian or Alaska, Asian, Native Hawaiian or Pacific Isander, Other). SCT denotes the score of the Sentence Comprehension Test. CRT denotes the score of the Cognitive Reflection Test. Additional socio-demographic controls in Column (2) and (4) include marital status, household income (low, middle, high), education (low, middle, high), employment dummy, location of residence type (suburb, countryside, omitted category: living in city). The behavioral variables B include individual's trust in fact checking (5-point answer scale: strongly distrust to strongly trust), general trust in people (5-point answer scale: strongly distrust to strongly trust), perception of fake news being a problem for the country (4-point answer scale: very big problem to not a problem at all), and political preference (omitted category: democratic, republican, independent, prefer not to answer). Media consumption I variables include the media consumption indicator and measures for the frequency of consumption of four different news sources: social media, television, online newspapers, printed newspapers; for each (5-point answer scale, not familiar with this source of news to the main source of news). The dependent variable (total score) is measured by the news-accuracy detection task. Participants assess the accuracy of 20 news items in the news-accuracy detection task—by answering the question: *"To the best of your knowledge, is this information in the above news item accurate?"*. The answer categories are *"Yes"* or *"No"*. For each correct answer, one point is received.

Table A6: Ability to detect Fake News (Socioeconomic and behavioral characteristics)

Figure A3: Participants' Engagement (word cloud)

Notes: At the end of the survey, participants receive the following final question: "Do you have any comments about our survey?"; with an open answer text field. This Figure shows a word cloud with all the important words used in sizes related to the frequency with which they are used.

Dependent variable: updating	
Treatment Sign × MTE	
positive × MTE	0.126**
	(0.06)
negative × MTE	-0.231***
	(0.07)
Treatment sign: positive	-0.156
	(0.15)
Treatment sign: negative	-0.148
	(0.16)
MTE	0.008
	(0.03)
Female	-0.155***
	(0.06)
middle aged	0.063
	(0.07)
old	0.085
	(0.08)
score sct	-0.006
	(0.04)
score crt	0.005
	(0.03)
single	-0.047
	(0.06)
currently employed	-0.087
	(0.06)
middle education	0.113*
	(0.07)
high education	0.060
	(0.08)
income middle	0.011
	(0.07)
income high	0.069
	(0.10)
living in suburbs	-0.043
	(0.06)
living in countryside	-0.042
	(0.08)
Controls Z_i	yes
N	2224
R^2	0.19

Notes: OLS estimates of specification (5.2). Standard errors are reported in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Socio-demographic controls X_i include gender, age (young, middle aged, old), and categories for ethnicity (Non-Hispanic White, Hispanic, Black or African American, American Indian or Alaska, Asian, Native Hawaiian or Pacific Isander, Other), SCT and CRT scores, marital status, household income (low, middle, high), education (low, middle, high), employment dummy, location of residence type (suburb, countryside, omitted category: living in city).

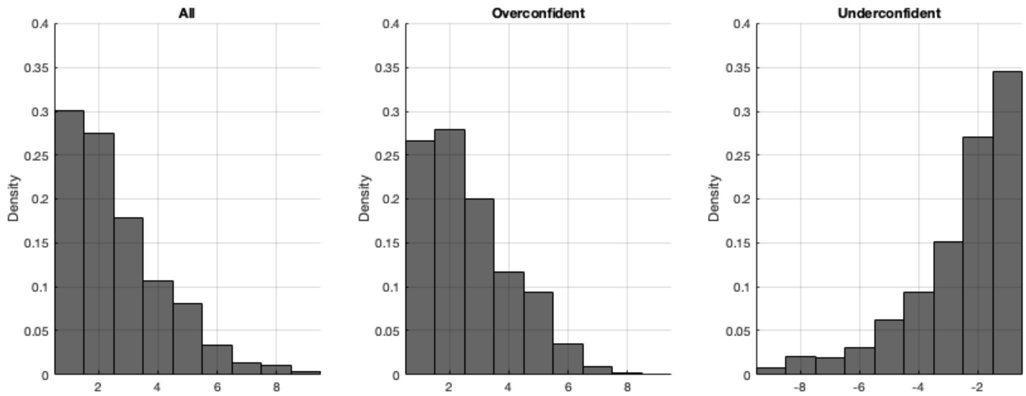
Table A7: Beliefs Updating and Feedback Sign (OLS estimates)

B Appendix: Additional Results & Robustness Checks

Prior Beliefs (pre-treatment) and the Belief–Ability Gap

The share of subjects with correct prior beliefs amounts to 13.47%, while 48.61% overestimate their news-accuracy detection ability and the remaining 37.92% underestimate it. Figure B1 plots the distribution of $error_i$ – i.e. the difference between the prior belief and the score in the first part of the task – for overconfident and underconfident subjects separately, as well as the the distribution of $error_i$ in absolute value for all subjects with incorrect prior beliefs. The average magnitude of the error in prior beliefs is 2.26 points, which implies a deviation from the true score of 37.79%. The error of overconfident subjects equals 2.64 points, while those of underconfident subjects is 2.57 points. This implies that the former believe their score is 42.42% higher than their true score, while the latter believe it is 45.75% lower.

Figure B1: Distribution of $error_i$ for mid-task beliefs. Left: absolute value of $error_i$ for all subjects with $error_i \neq 0$; center: $error_i > 0$ (overconfident subjects); right: $error_i < 0$ (underconfident subjects).



Robustness Check: General qualitative beliefs

We also explore the information treatment effect on general qualitative beliefs instead of quantitative beliefs. We elicited respondents' general qualitative beliefs about (i) their own and beliefs about (ii) the average American's ability to detect the accuracy of news items. Subjects can answer both questions on a scale from 1 (very bad) to 5 (very good). We ask these questions before and after the news-accuracy detection task, thus eliciting prior and posterior beliefs. Specifically, we ask subjects the following two questions:

1. *"In your opinion, how good is your ability to identify news or information that misrepresents reality or is even false?"*

2. *"In your opinion, how good is the average American's ability to identify news or information that misrepresents reality or is even false?"*

For each question, we construct a dummy variable equal to one if the subjects consider themselves (the average American) at least *"good"* in identifying news or information that misrepresents reality or is false and zero otherwise (elicited after the news accuracy detection task; i.e., the posterior qualitative beliefs).

We then run a logistic regression of this posterior qualitative beliefs dummy on $Treat_i^{Sign}$ that captures whether the individual i received in the score revelation treatment a positive or negative signal (i.e., information) about his ability and the vector of individual-specific characteristics X_i used in thebaseline specification of the paper. We also control for a dummy that equals one if subjects consider themselves at least *"good"* in identifying news or information that misrepresents reality or is false and zero otherwise (elicited before the news accuracy detection task; i.e., the post qualitative beliefs). Moreover, in this regression, we also control for individual i's performance (total score), effort, and finally for $MTESign_i$, subjects' "perception" of their ability ($MTESign_i$). Recall that $MTEsign_i$ captures whether subject i has had a positive, negative, or correct pre-treatment perception of his(her) ability. We define positive (negative) self-perception by comparing mid-task beliefs and the actual performance score during the first part of the news-accuracy detection task. A participant has a positive (negative) self-perception about his ability if he believes he has given more (less) correct answers than he actually did.

Table B1 presents the results. Panel A shows two key results. First, treated subjects significantly adjust their posterior qualitative ability beliefs according to the information they receive. For example, those who receive negative feedback are less likely to consider themselves "good" at identifying fake news (elicited after the news accuracy detection task i.e., the posterior qualitative beliefs). Second, those who receive negative feedback adjust their posterior qualitative beliefs about their own ability more strongly than those who receive positive feedback.

Panel B of Table B1 shows the results when considering the treatment effect on individuals' beliefs about the average American's ability to detect the accuracy of news items. We find that the information revelation treatment about one's own fake news detection ability does not affect subjects' beliefs about the average American's ability.

Dependent variable: Qualitative ability beliefs (posterior)		
	Panel A	Panel B
$Treat_i^{Sign}$		
positive	0.0347**	0.0127
	(0.0172)	(0.0200)
negative	-0.0550**	-0.0145
	(0.0263)	(0.0190)
Observations	2413	2413

Notes. Marginal effects of logistic regression. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Robust standard errors are in parentheses. Dependent variable: dummy variable equal to one if subjects consider themselves in panel A (the average American in panel B) at least "good" at identifying news or information that misrepresents reality or is false and zero otherwise (elicited after the news accuracy detection task; i.e., the posterior qualitative beliefs). $Treat_i^{Sign}$ that captures whether the individual i received in the score revelation treatment a positive or negative signal (i.e., information) about his ability. The specifications also control for socioeconomic variables: age, age squared, gender, income categories, education achieved, work status, residential area, ethnicity, and religion. In addition, we also control for individual i's performance (total score), effort, and finally for $MTESign_i$, subjects' "perception" of their ability ($MTESign_i$). Recall that $MTEsign_i$ captures whether the subject i has had a positive, negative, or correct pre-treatment perception of their ability. We define positive (negative) self-perception by comparing mid-task beliefs and the actual performance score during the first part of the news-accuracy detection task. A participant has a positive (negative) self-perception about his ability if he believes he has given more (less) correct answers than he actually did.

Table B1: Robustness Check: Qualitative Beliefs