WORKING PAPERS

N° 1519

Revised version October 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: October 10, 2024

Abstract

This paper quantifies the impact of a demand-side policy intervention on citizens' willingness to pay for protection against misinformation. We find that individuals generally lack proficiency in identifying fake news and overestimate their ability to distinguish between accurate and false content. Providing information-through-experience about personal susceptibility to fake news leads to belief updating and greater awareness of detection ability. Crucially, this increased awareness significantly raises individuals' willingness to pay for measures that protect against the harms of misinformation.

Keywords: Fake news, misinformation, personal susceptibility, experience, belief updating, willingness to pay, demand-side policy intervention, RCT experiments.

JEL Classifications: C83, D83, D84, D91.

We thank, without implicating, several colleagues who generously provided feedback on a preliminary draft of the paper. In particular, we are grateful to Fabrice Collard, Thomas Dohmen, John Duffy, Patrick Fève, Sergei Guriev, Emeric Henry, David Thesmar, Jean Tirole and Florian Zimmermann. 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, closely followed by polarisation within our societies."* This statement stresses the growing recognition of dis- and mis-information as a significant threat not only to societal cohesion, but also to economic stability.[1] Recent surveys reveal a paradox: while the average citizen recognizes the potential harms of fake news — 85% of Europeans and 82% of Americans consider it a problem for the country (Eurobarometer, 2018; Pew, 2019) — most also believe they are personally "immune" to being fooled by fake news. The large majority, 84% of Americans and 71% of Europeans, feel at least "somewhat confident" in their ability to detect fake news.[2] This disconnect suggests that citizens might have a misperception about their personal susceptibility to fall for fake news, which could undermine efforts to reduce the spread and impact of fake news.

Despite growing research on the effectiveness of supply-side interventions to reduce the spread of political false news (Guriev et al., 2023; Henry et al., 2022; Pennycook et al., 2021) and studies focused on profiling characteristics that affect citizens' ability to detect political fake news (Angelucci and Prat, 2024; Lyons et al., 2021; Pennycook and Rand, 2020a), a significant gap remains: how do individuals' beliefs about their vulnerability to fake news influence their economic behavior?

This paper addresses this gap by quantifying the impact of raising awareness about personal susceptibility to fake news on individuals' economic decisions. Citizens may differ not only in their detection ability but also in their awareness of their vulnerability to fake news. Underestimating one's susceptibility can lead to a false sense of security, amplifying the social and economic harms of fake news. This lack of awareness is particularly concerning within the broader economics of misinformation, where it can impede informed decision-making, distort market behavior, and undermine policy efforts aimed at mitigating misinformation's impact.

[1] See Assenza et al. (2024) for an analysis of how fake news influences business cycle fluctuations.

[2] 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.

Specifically, we investigate whether providing citizens with information about their ability to detect fake news influences their willingness to pay for protection against misinformation. In doing so, we shift the focus from supply-side to demand-side interventions, exploring the potential for self-awareness to drive proactive measures against misinformation. This novel approach distinguishes our research by examining the behavioral effects of updating beliefs about susceptibility, thus contributing a new dimension to the literature on misinformation economics.

We conducted a randomized controlled trial (RCT) with a representative sample of the US population, where participants were exposed to an incentivized fake news detection task. By providing half of the participants with feedback on their performance, we induce an exogenous shift in their beliefs about their ability to detect fake news. This allows us to test whether such belief updates lead to changes in their willingness to pay for a hypothetical product designed to mitigate the risk of being harmed by fake news, in the following we refer to it as the *"misinformation insurance"*.

By addressing the critical and unexplored question of how raising awareness of personal susceptibility to fake news influences economic decision-making, this paper makes several important contributions. To begin with, we expand the literature on the economics of misinformation by being the first to focus on demand-side solutions, shifting the emphasis from institutional regulation to individual action. Our results suggest that interventions targeting individual behavior, specifically by correcting misperceptions about personal susceptibility to misinformation, could play a crucial role in combating disinformation. This approach offers a feasible demand-side complement to supply-side regulatory efforts. Second, we contribute methodologically to the large literature using information provision experiments—by introducing a novel *information-through-experience approach* in our experimental design. In this setup, participants engage with a task that directly tests their ability to detect fake news, receiving feedback based on their actual performance. This immersive experience delivers information in a more personal and credible manner, thereby enhancing the relevance of the feedback. The resulting shift in participants' beliefs about their detection ability serves as a manipulation check, which we then use to investigate the causal effect of belief updates on economic behavior, such as their willingness to pay for protection against misinformation.

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.[3] Other studies have used similar quizzes

[3]We collect twenty news headlines from a set of fact-checked news items by different non-partisan orga-

to measure the ability of citizens to distinguish between accurate and fake news, in contrast to these studies, we do not focus exclusively on political fake news, but we cover a representative spectrum of topics and news outlets.[4]

Each headline contained a subheading highlighting the critical content of the news and the publication date. 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 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-through-experience 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; treated subjects 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, by comparing mid-task with posterior beliefs, we can analyze whether the information-through-experience treatment successfully raised awareness of susceptibility to falling for fake news, measured by changed ability 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. In the experiment, we give subjects (pre and post-experiment) a hypothetical $1000 budget, which they can allocate freely between consumption goods, health insurance, and misinformation insurance—an insurance

nizations (including Politifact, Snopes, Reuters, Science Feedback, and Factcheck.org). All news items appeared in the 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.

[4]While we are not interested in measuring individuals' actual "ability" to detect fake news, but rather on shifting their beliefs about their own detection skills, using a broad spectrum of topics remains important. Political science research indicates that men are more likely than women to engage in partisan political activities (Quaranta and Dotti Sani, 2018) and that women tend to express less interest in politics than men (Fraile and Gomez, 2017a,b). Hence, the average women will be less knowledgeable about political topics. It will not surprise that women's "ability" to detect political fake news is worse than men's. To mitigate this potential bias, we include a wide range of topics in our news items, such as crime, science, politics, the Covid-19 pandemic, climate change, and health. 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).

contract covering the risk of harm by fake news. This design enables us to establish a causal relationship between subjects' willingness to pay for "misinformation insurance" and the information provided about their personal susceptibility to fall for fake news. By comparing pre- and post-experiment choices, we can analyze how this information-through-experience treatment affects spending in relative terms.

We choose the label "misinformation insurance" for three key reasons. First, in the real world, fake news is becoming a widespread issue, leading major corporations like Bank of America and SAP to recommend best practices and business guidelines for mitigating the harm caused by misinformation. Additionally, companies such as JPMorgan already offer services that promise to protect clients from misinformation-related risks. These real-world examples make the concept of misinformation insurance both plausible and relevant.[5] Second, the concept of insurance inherently captures the notion of probability—while insurance reduces harm, it does not guarantee 100% protection from negative outcomes. Third, the framing "misinformation insurance" is more neutral compared to specific services like "fact-checking service". By framing the issue of misinformation in terms of risk management and incorporating it into an economic decision-making context, our paper provides a new approach to understanding the economic consequences of disinformation. This innovative concept not only provides a tangible metric for assessing the demand for anti-misinformation tools but also opens up new avenues for future research on market-based solutions to mitigate the impact of fake news. Our paper is the first to measure the general willingness to pay for protection against fake news.

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).

The key results of this paper can be summarized as follows. First, we find that the information-through-experience treatment significantly alters subjects' beliefs about their ability to detect fake

[5] Several consulting firms also offer services to counter and mitigate the negative effects of misinformation. For example, Greydient Analytics provides a "Disinformation Monitoring Service" to help businesses stay informed and protect themselves from disinformation, misinformation, and malinformation. Becker Digital, a veteran-owned consulting firm, works to protect communities and organizations from both misinformation and disinformation. KROLL assists clients in managing digital risks, including fake news and misinformation, while providing crisis response. **Links:** JP Morgan; Greydient Analytics; Becker Digital; KROLL; Bank of America; SAP.

news. This increased awareness of their susceptibility to fake news (i.e. belief updating) results in a more accurate self-assessment of personal susceptibility. Second, treated subjects who received a negative signal about their performance (i.e., those who had overestimated their ability to detect fake news) adjusted their ability beliefs downward. This shift in beliefs, induced by the information-through-experience treatment, in turn, causally increases subjects' willingness to pay to reduce the risk of being harmed by misinformation.

The academic literature still lacks consensus on which socioeconomic characteristics predict the ability to detect fake news (Arin et al., 2023). While this is not the focus of this paper, we provide systematic evidence on the effects of gender, age, and other individual-level characteristics on fake news detection ability. Unlike prior studies that focus exclusively on political fake news (e.g., Angelucci and Prat, 2024; Guriev et al., 2023; Pennycook and Rand, 2020a), the statements of a single politician (e.g., Barrera et al., 2020) or specific topics (e.g., Arechar et al., 2023; Lutzke et al., 2019), we cover a representative spectrum of topics and news outlets. We find that women, older respondents (above 54 years), and those who are single perform significantly better at 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 noted in Bago et al., 2020 and Pennycook and Rand, 2019) and the Sentence Comprehension Test. These results remain highly robust even after accounting for subjects' media consumption habits and political preferences. Notably, social media consumption emerges as a significant predictor: individuals who rely primarily on social media for news perform significantly worse.[6]

Our results are relevant from a policy perspective. The policy debate has explored a range of supply- and demand-side interventions to address the problem of fake news. Supply-side solutions aim to reduce the amount of misinformation or disinformation supplied into the market. These measures can be privately driven, such as social media platforms' content moderation policies, or state-led, through regulations requiring commercial information providers (e.g., social media platforms) to meet fact-checking obligations. Some researchers have already tested the effectiveness of tools in this area, including fact-checking, requiring confirmation clicks to access or share false content, or requiring comments before sharing (e.g., Pennycook and Rand, 2020b; Henry et al., 2022; Arechar et al., 2023; Guriev et al., 2023). In contrast, demand-side policies focus on enhancing citizens' ability to detect fake news. Such policies should include improving general educational

[6] 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 outperform men in detecting fake news. This difference is likely due to their focus on political news, an area where men tend to be more engaged, interested, and informed.

outcomes and providing specific training in media and digital literacy. These interventions would likely improve citizens' performance in detecting fake news but they require substantial financial investments and long-term planning.

While supply-side approaches address the problem at its roots, they have at least two notable drawbacks. First, the technology for creating fake news evolves rapidly, making it difficult for private and public actors to keep pace with those spreading fake news. Second, supply-side solutions risk becoming politicized. For example, in the United States, right-wing politicians have accused social media platforms of colluding with left-wing politicians and academic experts to censor the "truth". As Stiglitz and Kosenko (2024) highlights, "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)".

This paper leverages citizens' misperception of their personal susceptibility to fake news and proposes a simple, cost-effective demand-side solution. We implement a policy intervention that educates individuals about their personal vulnerability to fake news. Those who receive a negative feedback on their fake news detection performance significantly adjust their self-assessment of ability (beliefs)—and, most importantly, take action to protect themselves from the harms caused 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 construct our dataset by selecting twenty news headlines from a set of fact-checked news items. Each news items has been fact-checked by at least one of the non-partisan organizations including Politifact, Snopes, Reuters, Science Feedback, and Factcheck.org. In some cases, more than one organization fact-checked the same item. All selected news items appeared in the media between

May 2018 and May 2021.

Half of the news items in our dataset 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. Participants were not informed of this distribution, but were told that all news items had been fact-checked by independent third parties. By relying on established fact-checking organizations for both true and fake news, we ensured consistency and avoided unintentionally including false information presented as true. This precaution was essential to mitigate potential ethical concerns and secure IRB approval for the study. While not all real news is fact-checked, using fact-checked headlines helps maintain a rigorous standard. It is important to note that the specific selection of news headlines is not central to our research question. Our primary objective is to examine whether increased awareness of personal susceptibility to fake news affects the demand for protection against misinformation. The news quiz, therefore, functions as a manipulation check for participants' beliefs about their ability to detect fake news.

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. The twenty headlines we selected, including the publication date, publisher, topic, the news article's "highlights", and additional information are reported in Appendix Table D1.

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.[7] 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

[7] 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. Appendix C reports the entire questionnaire.

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., 2024; Roth and Wohlfart, 2020) to screen out distracted participants leading to potentially low-quality observations.[8] 75% of all respondents passed the attention check and are included in this study.[9] 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

[8] 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".

[9] 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)).

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 objectives are threefold. First, to quantify individual's susceptibility to fall for fake news and their beliefs about their own ability. Second, to test the effectiveness of a simple policy intervention using an *information-through-experience approach* to raise individual awareness of their susceptibility to fake news. Third, to examine how this intervention—providing participants with feedback on their ability to distinguish between true and fabricated information—affects 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) become aware of their personal susceptibility, and update their ability belief accordingly, leading to a more accurate self-assessment of their susceptibility to fake news.*

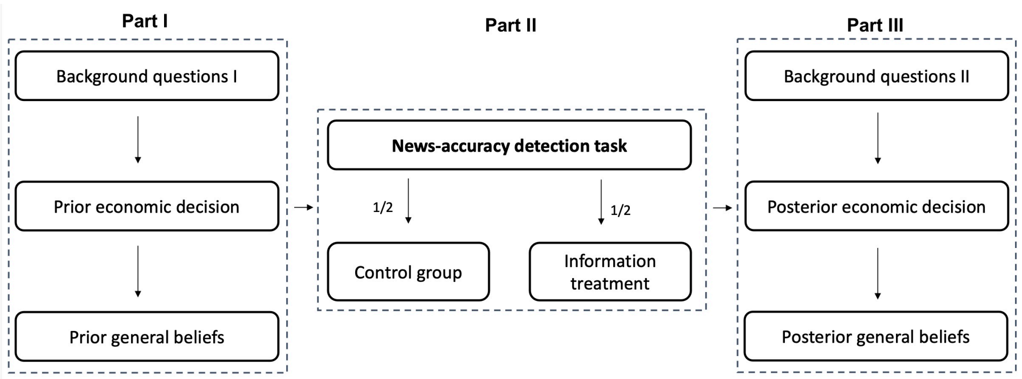
Hypothesis H2 (Economic choices): *Subjects receiving a negative signal, conditional on H1 being true, significantly increase their willingness to pay for misinformation insurance.*

Hypothesis H1 evaluates whether the information-through-experience treatment leads to a significant change in subjects' awareness of their personal susceptibility to fake news, measured through ability belief updating and accuracy. Hypothesis H2, assuming Hypothesis H1 holds, tests 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.

In addition to testing these hypotheses, we contribute methodologically to the broader literature on information provision experiments by introducing this *information-through-experience approach.* In this experimental design, participants perform a task that directly tests their ability to detect fake news and receive feedback based on their actual performance. This personalized signal contributes to enhance relevance and credibility of the information provided, making the feedback more impactful. The shift in participants' beliefs about their detection ability serves as a manipulation check, allowing us to assess the causal effect of belief updating on economic behavior.

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

Figure 1: Survey Structure and Timeline



- **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-through-experience related to their ability to detect the accuracy of news, after the first and before the second part of the task (Figure 3). In the case of the control group without any information provision, there is no signal. The task, the provided information-through-experience, 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.

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. This design allows the analysis of the effect of the treatment on spending in *relative* terms.

The label "misinformation insurance", we choose for three key reasons. First, the increasing prevalence of fake news has led major corporations, such as Bank of America and SAP, to adopt guidelines for mitigating misinformation risks, with companies like JPMorgan offering services to protect clients against such risks. These real-world practices make the concept of misinformation insurance relevant. Second, insurance inherently reflects the idea of risk mitigation, reducing harm but not guaranteeing full protection.[10] Third, the notion of a specific service such as "fact-checker service" could polarize. The framing "misinformation insurance" is neutral. 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-through-experience about their news accuracy detection performance. In other words, we measure the treatment effect of increased awareness on the economic choices by eliciting quantitative variations in spending in a

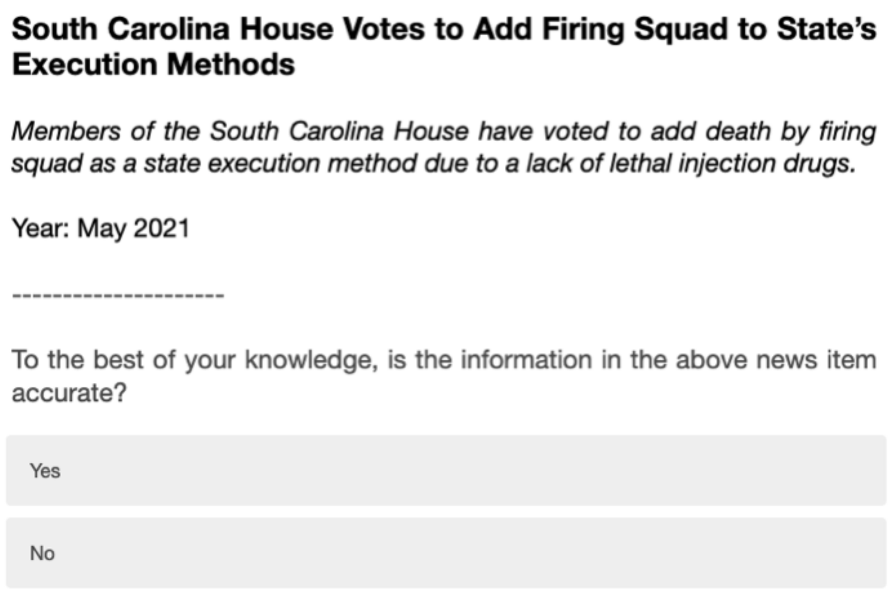
[10] A health insurance cannot protect you from getting cancer, but can reduce the probability of death and the costs associated with medical treatments. Similarly, a misinformation insurance (a service like a fact-checker) can reduce the risks of being harmed by fake news. A fact-checking service provides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

"hypothetical" scenario.[11]

3.3 Information-through-experience teatment 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 of truth individuals assign to the news items.[12] 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.[13]

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



[11]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 work well to predict behavioral changes and estimate population aggregates, which is the objective of our study.

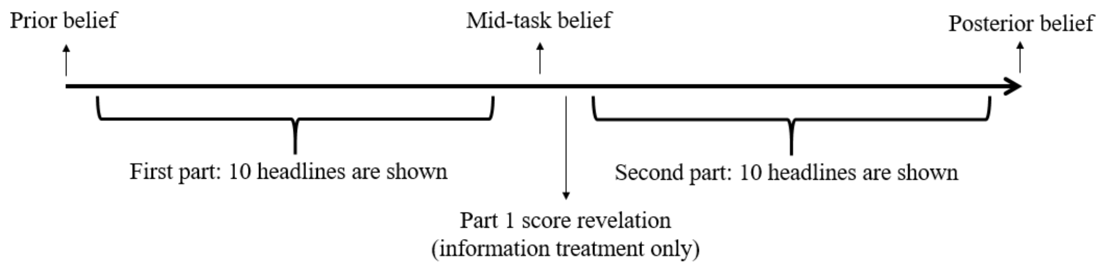
[12]This approach differs from Angelucci and Prat (2024), who study relative probabilities by showing six news items on the same screen and informing the participants that 50% of them are true.

[13]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.

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.

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.[14] At the end of the experiment, two participants will be randomly selected and their total scores will form the basis of the donation.

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



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

[14]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.

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"*.

Treatment. We randomly allocate 50% of the subjects to our *information-through-experience treatment*. Treated subjects receive information on their score after the first part of the task (i.e., 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).

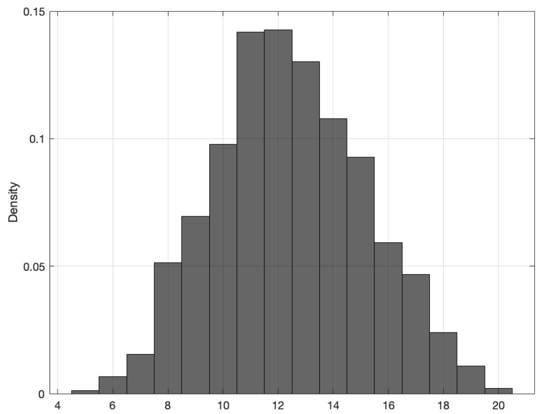
Potential Design Concerns. We reduced the potential for experimenter demand effects by choosing a between subject design compared to a within subject design (e.g., Charness et al. (2012) and Zizzo (2010)). In addition, our experimental design is characterized by three central features that the literature 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.[15]

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.[16] 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).[17]

[15]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 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 ($p = 0.301$).

[16]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."*

[17]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 (as e.g. in Anderson et al., 2012)

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.*

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).[18] 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.[19] 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).

[18]Table 1 only reports significant results. Appendix Table A6 shows the full estimation results.
[19]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)

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 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.[20]

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.[21] 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

[20] 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.

[21] 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).

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.[22] While Sindermann 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).[23]

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[24] 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).[25]

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

[22]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 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.

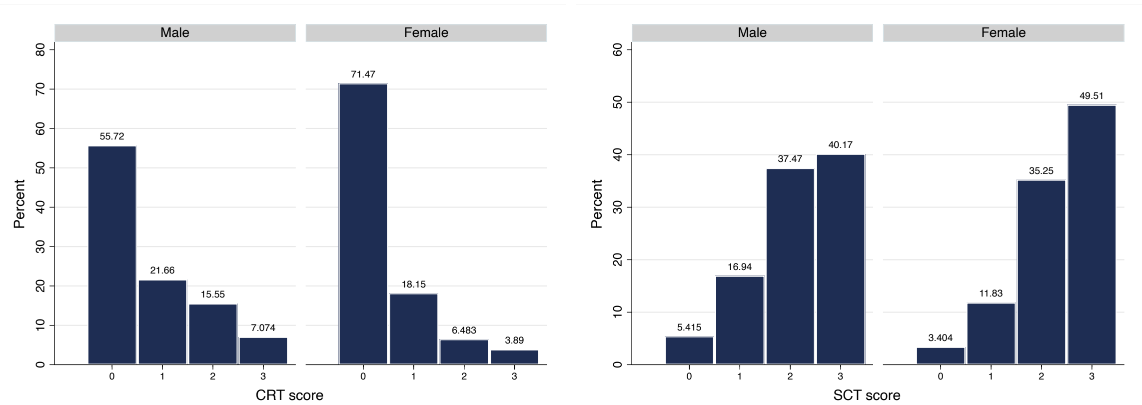
[23]Appendix Figure A2 shows the distribution of the performance score by gender, the gender differences in the mean ($t = -2.3054$, $p = 0.0212$), median (MWU-test, $z = -2.742$, $p = 0.0061$), and the distribution (Kolmogorov-Smirnov test) are statistically significant.

[24]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.

[25]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 e.g. Cueva et al. (2016); Pennycook et al. (2016); Ring et al. (2016); Toplak et al. (2014)

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 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.[26] We conclude that the survey experiment successfully engaged the participants. Appendix Table A5 shows that the sample is well-balanced between control and treatment groups, across various socioeconomic dimensions such as gender, age, race, income, education, and employment status.

This section explores the potential impact of a simple policy intervention— aimed at increasing participants' awareness of their actual susceptibility to fake news—on their ability belief updating, and their willingness to pay for misinformation insurance. In particular, we test the two hypotheses described in Section 3, evaluating whether raising awareness through this intervention leads to a significant change in beliefs and economic behavior.

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.

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

5.1 Treatment Effect on Beliefs

This section investigates whether and to what extent subjects' beliefs about their ability respond to information-through-experience they receive. Treated subjects are provided with feedback on their fake news detection score after completing the first and before starting the second part of the task (see Figure 3). For the analysis, we focus on 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, as shown in Figure 3.

Given participants' lack of prior experience with the task, concerns may arise regarding the accuracy of their initial beliefs. To overcome this potential issue, we focus on the mid-task beliefs—elicited after subjects have assessed the first 10 headlines—allowing them to become familiar with the task and reducing uncertainty associated with their initial judgments. Hence, our analysis compares mid-task beliefs (formed after exposure to the task) with posterior beliefs, elicited after the entire task is completed.[27]

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 information-through-experience 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; $POE_i = |belief_i^{posterior} - score_i^{part2}|$. $Treat_i$ is the variable of interest, the treatment dummy that equals one if the individual i received the information-through-experience 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

[27] The results of this paper remain robust when using prior beliefs instead of the mid-task beliefs. These results are available upon request.

[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.

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 of subjects to treatment and control groups allows us to conclude that providing participants with information-through-experience about their fake news detection ability causally enhances their awareness of their personal susceptibility by improving the accuracy of their self-assessed ability beliefs. 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.

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 the 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)

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) 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.

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 that identifies 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 the 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-through-experience they receive. The information (signal) increases awareness of personal susceptibility to fake news. As a result, 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-through-experience 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 information-through-experience treatment a positive or negative signal (i.e., information) about his ability.[29]

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

[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.

[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

magnitude of the treatment matters for the magnitude of belief updating. The adjustment in beliefs, resulting from the information-through-experience 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 information-through-experience treatment than participants who received the information that underestimated their ability (i.e., positive feedback). 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.*

[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.

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 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-through-experience, we measure subjects' economic choices. Subjects must allocate their budget between consumption goods, health insurance, and misinformation insurance, 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 information-through-experience 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

[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.

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 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: $\Delta wtp_i = \beta_0+\beta_1 Treat_i \times updating_i+\beta_2'\boldsymbol{X_i}+\beta_3'\boldsymbol{\tilde{Z}_i}+\epsilon_i$. 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

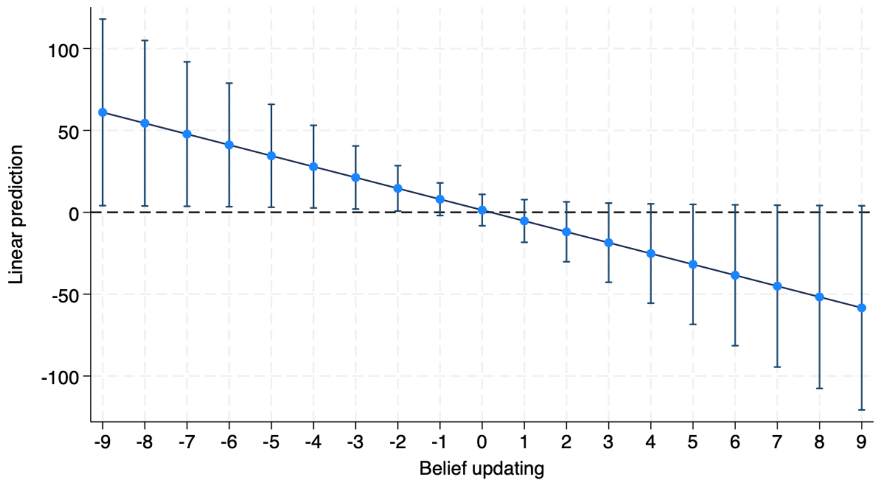
Note that the sign and magnitude of this average treatment effect reported in Table 4 are not very informative. The previous Section showed that the direction (Result 3) and magnitude (Result 4) of the information-through-experience effect on beliefs depend on whether the subject received positive or negative feedback concerning his ability to detect fake news. Treated subjects significantly adjust their ability beliefs according to the information-through-experience 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

[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.

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.

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 sub-

[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).

jects 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 poses significant risks to individuals, society, and the economy. This paper sheds light on the crucial role of individual awareness in shaping behavior and mitigating the spread of misinformation. By empirically assessing how belief updating affects citizens' willingness to pay for protective measures, we contribute to a more comprehensive understanding of the economic implications of fake news. Our findings provide evidence for the effectiveness of straightforward demand-side policy interventions that raise awareness about personal susceptibility to fake news, offering new insights for both policymakers and researchers.

The debate around interventions to counter fake news has largely focused on regulation, such as urging social media platforms to moderate and verify content. However, the rapid development of AI presents challenges for existing detection tools, which may struggle to remain accurate and reliable.[38] Suggested demand-side interventions include improving general educational outcomes or providing specific media literacy training, both of which would improve citizens' ability to detect fake news. However, these solutions require significant financial investment and long-term planning.

This paper is the first to empirically assess the effectiveness of a low-cost demand-side policy intervention that educates citizens, through direct experience, about their own vulnerability to fake news. Through a novel *information-through-experience* approach, we offer personalized feedback on participants' actual ability to identify fake news. This feedback not only helps gauge individual susceptibility, but also serves as a manipulation check to explore how belief updates influence behaviors, such as willingness to pay for misinformation protection.

[38] A further challenge to this approach is the potential for political tensions. For example, US-based tech companies' efforts to fact-check and curb misinformation have faced criticism from right-wing politicians who accuse them of colluding with governments and academics to censor conservative views.

Our findings show that participants who receive negative feedback significantly revise their self-assessment and show a greater willingness to pay for protection against misinformation. They emphasize the importance of understanding people's self-assessed ability to detect misinformation and the benefits of interventions that improve awareness of their limitations. By enhancing self-assessment accuracy, we empower individuals to make more informed decisions and reduce misinformation's harmful effects. We show that simple policy interventions can effectively raise awareness of vulnerability to fake news, which will be crucial in curbing its spread and protecting society.

Bibliography

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

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*.

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).

Assenza, T., Collard, F., Feve, P., and Huber, S. (2024). From Buzz to Bust: How Fake News Shapes the Business Cycle. *CEPR Discussion Paper No. 18912*.

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.

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.

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.

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

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.

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

Faia, E., Fuster, A., Pezone, V., and Zafar, B. (2024). Biases in information selection and processing: Survey evidence from the pandemic. *The Review of Economics and Statistics*, 106(3):829–847.

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.

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*.

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).

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.

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.

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.

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.

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